Biochemical Analysis of the Site Specific Recombinases of Yeasts and of Bacteriophage P1

Thesis

UNRESTRICTED

Biochemical Analysis of the Site Specific Recombinases of Yeasts and of Bacteriophage P1

Leonie Helen Ringrose

A thesis submitted in partial fulfilment of the requirements of the Open University for the degree of Doctor of philosophy.

July 1997

Sponsoring establishment
National Institute for Medical Research, London

Collaborating establishment
European Molecular Biology Laboratory, Heidelberg

Date of award : 1 December 1998

Some of the results presented here appear in the following publications:

Buchholz, F., Ringrose, L., Angrand, P.-O., Rossi, F. and Stewart, A.F.
Different thermostabilities of FLP and Cre recombinases: implications for applied site specific recombination.
Nucleic Acids Research, (1996) **24**, 4256- 4262

Ringrose, L., Angrand, P.-O. and Stewart, A.F.
The Kw recombinase, a new integrase from the yeast *Kluyveromyces waltii*.
European Journal of Biochemistry, (1997) **248**, 903-912.

Ringrose, L., Lounnas, V., Ehrlich, L., Buchholz, F., Wade, R. and Stewart, A.F.
Comparative kinetic analysis of FLP and Cre recombinases: Mathematical models for DNA binding and recombination.
Journal of Molecular Biology, (1998) in press.

This thesis is dedicated to myself,
without whom none of it would have been possible,
or necessary.

Table of Contents Page

List of Tables Page

Abbreviations

ATP	Adenosine triphosphate
bp	base pairs of DNA
BSA	Bovine Serum Albumin
Ci	Curie
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
dTTP	Deoxythymidine triphosphate
DEPC	Diethyl Pyrocarbonate
DTT	Dithiothretiol
EDTA	Ethylenediaminetetraacetic Acid
EMBL	European Molecular Biology Laboratory
kb	Kilobase pairs of DNA
kDa	Kilodalton
FLP	The site specific recombinase encoded by the 2 micron plasmid of the yeast Saccharomyces cerevisiae
FRT	The FLP recognition target DNA sequence
Kw	The site specific recombinase encoded by the 2 micron like plasmid of the yeast Kluyveromyces waltii
KRT	The Kw recognition target DNA sequence
Cre	The site specific recombinase encoded by bacteriophage P1
LoxP	The Cre recognition target DNA sequence
PEG	Polyethylene glycol

TEMED	N,N,N',N'- Tetramethylethylenediamine
X-gal	5-Bromo-4-chloro-3-indolyl-β -D-galagtopyranoside.

Abstract

The site specific recombinases FLP from *Saccharomyces cerevisiae* and Cre from bacteriophoge P1 are finding increasing use in reverse genetics. This thesis presents an investigation and comparison of the biochemical properties of site specific recombinases, with a view to optimising their use in genomic manipulation strategies.

Chapter 3 describes the characterisation of a novel site specific recombinase, the "Kw" recombinase, from the yeast *Kluyveromyces waltii.* The results show that Kw shares features with FLP and Cre recombinases which make it potentially useful as a tool for genomic manipulation.

In Chapter 4, kinetic properties of FLP and Cre are compared, showing that Cre has a higher affinity for the LoxP site than FLP for the FRT. The kinetics of the full excision reaction are compared for FLP and Cre. A mathematical modelling approach is used in combination with measured DNA binding rates, to determine other parameters which best describe the recombination behaviour of each recombinase. This analysis suggests that the synaptic complex is more stable for Cre than for FLP.

In Chapter 5, results show that FLP recombinase is thermolabile in an *in vitro* recombination assay, whereas Cre is thermostable, and that this difference also affects DNA binding. In addition, two FLP mutants with altered thermostabilities are described.

In Chapter 6, the effect of distance between FRT sites on FLP mediated excision *in vitro* is examined, showing that FLP is unable to excise a substrate with 74 base pairs between its target sites, and that the optimum distance, giving the fastest initial rate of recombination, is between 400 and 700 base pairs. At a given protein to substrate ratio, recombination efficiency is greatly decreased at distances of 8kb and above.

Foreword

The results presented in this thesis fall into four separate chapters, each of which forms an independent unit. For this reason, each chapter has a separate discussion, in which the interpretation of data, the contribution made by the work, and the further work to be done are discussed in detail. These points are not therefore covered by the final summary.

CHAPTER 1.
INTRODUCTION

1.1 Genetic Recombination

Genetic recombination occurs in the DNA of all organisms, and can be divided into three classes: 1) homologous or general recombination; 2) transpositional recombination and 3) conservative site -specific recombination. Each of these types of recombination was originally discovered from genetic studies. Later biochemical analysis has revealed details of the reaction mechanisms and the proteins involved, many of which have in turn been harnessed for application to reverse genetics, the manipulation of living genomes.

1.1.1 General recombination

General, or homologous recombination, which involves the reciprocal exchange of DNA between two homologous molecules, is fundamental to all organisms, both as a source of genetic diversity and for the repair of damaged DNA. Homologous recombination was first described in 1911 (Morgan) in *Drosophila*. Subsequently, genetic maps derived from crossover frequencies revealed that genes are arranged in a fixed linear order, a concept held to be universal until some 40 years later. In the last few decades, the study of homologous recombination in *E.coli* has lead to the identification and isolation of most of the enzymes required. The central enzyme is the RecA protein, which promotes homologous pairing and strand exchange. In addition to RecA, 14 other *E.coli* gene products are required, as well as the enzymes involved in general DNA metabolism (West, 1994). The initiation and homologous pairing steps of the reaction have been reconstituted in vitro (Kowalczykowski 1994). Recently, RecA homologues have been identified in several other bacterial species, in yeasts, a filamentous fungus, flies, mouse, chicken and human, suggesting common mechanistic features throughout evolution (West 1994, Dunderdale and West 1994, Calero et al. 1994, Schmitt et al. 1995, McKee et al. 1996).

Homologous recombination has proved an invaluable tool for the genetic manipulation of bacteria, yeast and mammalian cells in

culture, where foreign DNA with partial homology to a chosen target sequence can be stably integrated into the genome.

1.1.2 Transpositional recombination

Transpositional recombination, the process by which transposable genetic elements (transposons) "jump" from one place to another in the genome, was first described in maize by McClintock (1951). The concept of movable DNA elements challenged the dogma of the time, but it was not until some 20 years later that transposons themselves were identified in prokaryotes. Transposons have since been found in all organisms studied, often representing 1% - 10% of the genome. Transposons encode the proteins and DNA sequences necessary for their own maintenance, and usually confer no selective advantage on their host (however some prokaryotic transposons encode antibiotic resistance or pathogenicity). For this reason they are generally considered to be "DNA parasites." Transposition mechanisms and the proteins responsible have been well characterised for several transposons (for reviews see Sherratt, 1991 and Sadowski, 1993) and fall into two broad classes. In replicative transposition, a copy of the integrated transposon is inserted at another site in the genome. This class includes the retrotransposons, which are duplicated via a cDNA copy of an RNA transcript. Many retrotransposons also resemble retroviruses in the organisation of their genes. The second mechanism of transposition is "cut and paste", in which the entire element is excised and "hops" to another site. Most transposable elements can jump or be copied to many sites in the genome, and usually short duplications are made by DNA repair at the site of insertion. In some cases the entire transposon may be duplicated.

The advantages of movable genetic elements have not been overlooked by the field of reverse genetics. For example, in *Drosphila* , random insertion of foreign DNA into the genome is achieved using derivatives of the "P-element" transposon, which occurs naturally in the genome of some Drosophila strains (Rubin and Spradling 1982). Integration of foreign DNA into plant genomes (*Arabidopsis* and Tobacco) is achieved by *Agrobacterium* mediated

delivery of modified T-DNA transposons (Hernalsteens et al. 1992, Van-Sluys et al., 1987). One drawback of using transposable elements to integrate foreign DNA is that targetting is not possible.

1.1.3 Conservative site specific recombination

Conservative site specific recombination occurs at precise target sites, resulting in reciprocal exchange of DNA with no loss or gain of nucleotides. Site specific recombinase enzymes have been likened to "a biological fusion and streamlining of the reactions carried out by restriction enzymes and DNA ligase" (Qian et al., 1990).

Conservative site specific recombination was first described by Campbell (1962) for the integration of bacteriophage λ into the *E.coli* chromosome, and has since been described in yeasts, bacteria and other bacteriophages (for reviews, see Stark et al., 1992 and Sadowski, 1993). The *in vivo* roles of this type of recombination are varied, including phage integration and excision, switching of gene expression, plasmid copy number regulation, and resolution of multimeric plasmids. Site specific recombinases are also often encoded by transposons where they are required for resolution of transposition intermediates.

Accurate *in vitro* site specific recombination was first described for the bacteriophage λ system (Nash, 1975). This and other subsequent *in vitro* systems for site specific recombination have led to the purification of many of the proteins involved, and extensive characterisation of their modes of action. On the basis of certain features of the reaction mechanism and of amino acid sequence comparison, the site specific recombinases so far identified fall into two families: the integrase family, and the resolvase-invertase family (reviewed below).

Some site specific recombinases are finding increasingly widespread use as tools for *in vivo* genetic manipulation of bacteria, yeasts, plants, flies and mammals (reviewed in Kilby et al. 1993). This approach has the advantage that recombination is site

specific, however the target sites for the recombinase first have to be introduced into the genome by homologous or transposon mediated recombination.

1.1.4 Other types of recombination

This section has focused on recombination systems which are well characterised biochemically. Other types of recombination exist which have yet to yield to *in vitro* analysis. One form of site specific recombination whose mechanism has recently begun to emerge is V(D)J recombination, which generates diversity in the vertebrate immune system. During differentiation of B-lymphocytes, immunoglobulin light and heavy chain genes consisting of different combinations of variable and constant regions, are assembled from multiple coding sequences. T- cell receptors are assembled by a similar process (reviewed in Gellert, 1992a and b).

Recombination takes place at specific target sequences flanking the recombined regions. Both conservative and non conservative junctions are formed: The "signal joint" is made in a conservative manner, whilst the "coding joint" is made with an imprecision which further contributes to diversity. Until recently the mechanism of recombination was obscure, but the past two years have seen the cloning of a number of the proteins involved, and the establishment of an *in vitro* system in which the initial cleavage steps can be accurately carried out (for review, see Oettinger, 1996). Interestingly, several of the proteins so far analysed have similarity to microbial proteins. Murine VDJP, which carries out site-specific joining *in vitro*, shares a region of homology with bacterial ligases (Halligan et. al 1995). RAG I and RAG II, which can carry out the initial cleavage events *in vitro*, bear similarity, respectively, to integrase family site specific recombinases, and integration host factor (a bacterial accessory factor involved in several recombination reactions) (Bernsteln et al. 1996). Despite these apparent similarities, the emerging picture of the molecular mechanism of VDJ recombination indicates that this represents an entirely new class of recombinases.

1.2 Conservative Site Specific Recombination

The following section will focus on conservative site specific recombinases of the integrase and resolvase/invertase families.

1.2.1 Two familes share common features

The DNA target sites used by both families vary in length and complexity but in all systems, the site of recombination contains two identical or near identical repeat elements of 9 to 13 base pairs in inverted orientation to each other, separated by a "spacer" of 6 to 8 base pairs (integrase family) or 2 base pairs (resolvase/ invertase family). Recombination occurs between two such inverted repeat sites. The repeat elements are bound by recombinase, which catalyses strand breakage, exchange and rejoining in the spacer region. During recombination, a covalent protein-DNA intermediate is formed between the phosphate at the point of DNA cleavage and the catalytic amino acid of the recombinase. In the case of the integrase family, the catalytic amino acid is a conserved tyrosine, which forms a covalent link to the 3´ phosphate at the point of strand cleavage. In the case of the resolvase/ invertase family however, the catalytic amino acid is a conserved serine, which attaches to the 5´ phosphoryl end of the broken DNA strand. In both families, the phosphoprotein bond is only broken upon religation of the DNA strands. Topoisomerases also proceed via a covalent protein/ DNA intermediate. This intermediate is thought to conserve the energy in the phosphodiester bond, as no outside energy source (such as ATP) is required. For reviews of this topic, see Stark et al. (1992), Sadowski (1993), Sadowski (1995) and Jayaram, (1994).

1.2.2 DNA target sites and accessory proteins

Many recombinase target sites contain additional binding sites for the recombinase itself, and for accessory proteins.

In the **integrase family,** the basic inverted repeat unit consists of two recombinase binding sites (often called "half sites") separated by a non palindromic spacer of 6 to 8 base pairs. Built around this basic unit, various levels of complexity are seen. The simplest type of site, the inverted repeat alone, is exemplified by the Lox P site, the target of Cre recombinase of bacteriophage P1 (Hoess et al., 1984, Hoess and Abremski, 1984). Cre is normally responsible for resolving dimers of the P1 genome and thus ensuring correct segregation of the phage upon cell division (Sternberg and Hamilton, 1981). More rarely, Cre integrates the P1 genome into the bacterial chromosome at sites called LoxB, (Sternberg et al., 1981). Cre is the only protein required for recombination at Lox P and LoxB (Abremski and Hoess, 1984) .

A slightly more complex situation is represented by target sites which contain additional binding sites for the recombinase protein. This is the case in the yeast subfamily of integrase class recombinases, defined by Utatsu et al. (1987). The best characterised member of this family is the FLP recombinase, encoded on the 2μ plasmid of *S. cerevisiae* . FLP recognises sites on 2μ called "FRT" (FLP recombination target), and is responsible for maintaining the copy number of 2μ (reviewed by Sadowski, 1995). Other yeast family recombinases and their putative target sites have been identified by sequence comparison of 2μ with the 2μ- like plasmids of other yeast species (Utatsu et al. 1987, Chen et al. 1986, Chen at al 1992(a)). These target sites all contain the basic inverted repeat element, but there may be between one and four adjacent additional direct repeat elements. Two yeast family recombinases have so far been characterised *in vitro*: FLP and the "R" recombinase from *Zygosaccharomyces rouxii* (Araki et al. 1992). In both cases each element is bound by recombinase, but only the inverted repeat is required for recombination, and no accessory proteins are necessary (Senecoff et al. 1985, Andrews et al. 1987, Araki et al. 1992, Sadowski 1995).

The most complex type of site, containing the inverted repeat element plus several binding sites for accessory proteins, is represented by the "attP" site of bacteriophage λ. The λ integrase, from which the integrase family gets its name, catalyses

integration of the phage genome into that of its host by recombination between the attP site on the phage, and the attB site on the bacterial chromosome. Whilst the att B site is a simple inverted repeat of 25 base pairs, the attP site covers 240 base pairs, containing, in addition to the 25 base pair inverted repeat, 11 other protein binding sites. Five of these sites, called "arm" sites, are bound by the N- terminal region of λ integrase itself (the C-terminal region binds the inverted repeat and carries out catalysis). The remaining six sites are specific binding sites for accessory proteins. There are three sites for host encoded IHF (integration host factor) which is required for both integration and excision. Two sites are bound by phage encoded Xis which is required only for excision. Finally, there is one binding site for host encoded Fis protein, which is not required for either reaction but stimulates excision when Xis is in short supply. (Reviewed by Landy, 1993).

In the integrase family, both halves of the inverted repeat site are usually bound by the same recombinase. An interesting exception is the XerC/ XerD system in *E.coli*, in which two related recombinases are required for recombination, each of which specifically recognises one half of the recombination target site. (Other accessory proteins are also required) (Blakely et al. 1993, for review see Sherratt et. al., 1995).

In summary the integrase family shows several levels of complexity in terms of target site structure and requirement for accessory proteins. Amongst the members of the resolvase/ invertase family so far characterised, the situation is rather more simple.

Resolvases carry out resolution of transposition intermediates by excision of DNA between two target sites (Examples are the Tn3 and γδ resolvases). They have DNA target sites consisting of an inverted repeat with a spacer of 2 base pairs. Adjacent to this site are two more inverted repeats, which are also bound by protein but where strand exchange does not take place. This is somewhat reminiscent of the yeast family target sites described above, however the resolvase "sub sites" are required for recombination, where they are

necessary for correct alignment of the recombination sites. No accessory proteins are required.

Invertases catalyse inversion of DNA segments between two target sites (examples are the Gin invertase, which changes the host range of phage mu, and the Hin invertase, which changes the antigenicity of *Salmonella*.) For these recombinases, the inverted repeat target site is flanked by a recombination enhancer sequence containing binding sites for the Fis protein, which is required for the inversion reaction. (For reviews of resolvases and invertases, see Sadowski (1993), Stark et al. (1992) and Johnson (1991).

1.2.3 Amino acid sequence similarity

The resolvase/invertase family of site specific recombinases comprises 44 members (Sadowski 1995). Amino acid sequence similarity amongst groups within this family are high: the invertases are highly related, (about 60% amino acid identity, Craig, 1988). Within the resolvase family, amino acid sequence identity is about 30%, and further subgroups are identifiable whose members are functionally interchangeable and highly related (about 80% amino acid identity, Craig, 1988) .

The integrase family of site specific recombinases comprises about 30 members which have been identified by amino acid sequence comparison (Argos et al., 1986, Abremski and Hoess, 1992). The family is not very highly conserved, but has some characteristic homologies, which are summarised in Figure 1.1 and Table 1.1 The most striking feature of the family is the absolute conservation of four amino acids. In FLP, these are arginine 191, histidine 305, arginine 308 and tyrosine 343. For λ, FLP and R, the conserved tyrosine becomes covalently attached to the 3′ phosphate during catalysis. (Pargellis et al. 1988, Evans et al. 1990, Chen et al. 1992 (b)). Of the other three conserved residues, arginine 191 of FLP and the corresponding arginine of R (Chen et al. 1992 (b) has been shown to play a role both in cleavage (Parsons et al. 1988) and in strand ligation (Friesen and Sadowski 1992). Histidine 305 is also necessary for strand ligation by FLP, whilst arginine 308 appears to play an accessory role in cleavage (Parsons et al. 1988). For Cre, the

individual roles of the four conserved residues have not been studied directly, but mutational analysis of Cre shows that a change in any one of the four invariant residues gives a protein which is defective in catalysis but unaffected in DNA binding (Weirzbicki et al. 1987). The role of arginine 173 of Cre in catalysis was confirmed by Abremski and Hoess (1992). The absolute conservation of these four residues throughout the integrase family suggests that these proteins share a common catalytic mechanism.

The nonalignable N- terminal region of λ integrase (thin line in Figure 1.1) contains the domain which binds the attP "arm" sites whereas the alignable C- terminal domain binds the inverted repeat recombination target. The analysis of Argos et al. (1986), based on seven recombinase sequences, identified the "integrase box", a stretch of 40 amino acids, with 10 to 68% identity (see Table 1.1) spanning residues 305 to 345 of FLP (corresponding to 308 to 343 of λ). This "signature" enabled the identification of many more members of the integrase family, 28 of which served as a basis for later analysis by Abremski and Hoess (1992), who used methods for detecting local regions of homology in unaligned but related sequences. This approach identified a second region of homology ("box I" on Figure 1.1) spanning amino acids 209 to 223 of λ .

This region corresponds to box I of the yeast subfamily (amino acids 185 to 203 of FLP), which had previously been identified by Utatsu et al. (1987) by comparison of putative recombinase coding regions from yeast 2μ like plasmids. The yeast subfamily is more highly conserved than the rest of the integrase family, containing, in addition to box I (about 50% to 60 % identity) a second highly conserved region (box II, amino acids 295 to 315 of FLP) which is about 60% to 70 % identical. As seen in Figure 1, these two boxes contain three of the four invariant catalytic residues.

In summary, although the site specific recombinases of the integrase family share limited amino acid sequence homology, they appear to use a common catalytic mechanism to recombine their distinct DNA target sites. In the case of the yeast subfamily, inconsistencies in the relatedness of recombinase amino acid sequences when compared to the relateness of 2μ -like plasmids as

Figure 1.1 Similarity in the integrase family of site specific recombinases. *Open boxes* indicate regions of proteins which can be aligned (Argos et al., 1986, Abremski and Hoess, 1992, Utatsu et al., 1987). *Single lines* indicate regions of λ and Cre unable to be aligned with each other or with other integrase family recombinases (amino acids 1-144 and 1-102 respectively). Coloured regions show stretches of highest similarity. The four absolutely conserved residues are indicated as R, H, R and Y, corresponding to residues 191, 305, 308, and 343 of FLP. (residues 212, 308, 311 and 342 of λ, and residues 173, 289, 292 and 324 of Cre). *Yellow region*: the "integrase box" (Argos et al., 1986) corresponds to amino acids 308- 343, 289-326 and 305 -345 for λ, Cre and FLP respectively, and contains 10 to 21% identical amino acids for pairs of these recombinases (Table 1.1). In the yeast subfamily, amino acid identity within the integrase box is above 60% (Table 1.1). *Red regions*: Yeast boxes I and II (Utatsu et al., 1987; residues 185-203 and 295-315 of FLP) are 55 to 75% identical for the yeast family recombinases (Table 1.1). *Pink regions:* The integrase box I (Abremski and Hoess, 1992; residues 209-223 and 170 to 184 of λ and Cre respectively) is 13 to 20% identical for λ, Cre and FLP (Table 1.1). *Orange regions:* overlap between yeast box II and integrase box.
The proteolytic fragments of FLP and Cre which have specific DNA binding activity (Pan et al., 1991, Chen et al., 1991, Hoess et al., 1990) are flanked by *black arrows*.
The catalytic domain of λ integrase, spanning amino acids 170 to 356) whose crystal structure has recently been determined, is flanked by *asterisks* (*) (Kwon et al., 1997). The stretch of amino acids (residues 279 to 329) which contains 12 of the 13 amino acids which determine the DNA binding specificity of λ integrase are shown by a *horizontal grey line* (Yagil et al., 1995, Dorgai et al., 1995).

Table 1.1 Amino acid sequence identity in members of the integrase family of site specific recombinases. The percent amino acid identity between pairs of recombinases is given for the regions shown in figure 1.1

Figure 1.1

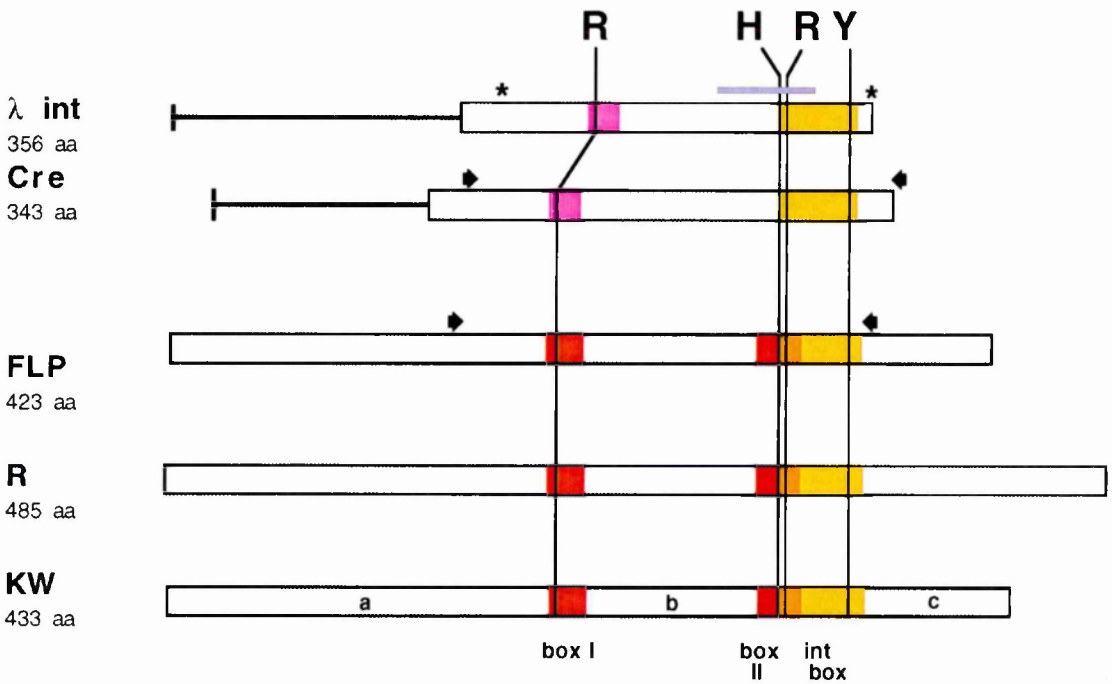


Table 1.1

	a	box I	b	box II	int box	c
λ:cre		20			18	
λ:FLP		13			21	
cre:FLP		20			10	
FLP:R	21	56	36	60	60	30
FLP:KW	17	55	39	75	68	26
KW:R	24	61	40	55	63	34

a whole, have led to the suggestion that these recombinases may have coevolved with their target sites (Murray et al., 1988).

1.2.5 The outcome of recombination: integrase family

All integrase family recombinases so far characterised carry out recombination via a Holliday junction intermediate, formed by the cleavage and exchange, at a specific nucleotide position, of one pair of DNA strands at one end of the spacer. The existence of Holliday junction intermediates has been shown for λ integrase (Nunes-Düby et al., 1987, Kitts and Nash, 1987) FLP (Meyer-Leon et al., 1988) and Cre (Hoess et al., 1987). To make a recombinant product, the crossover point moves along the spacer to the other end, where the Holliday junction is resolved by a second pair of cleavage and ligation events which are essentially the chemical reverse of the first exchange step. A short stretch of heteroduplex DNA is left.

The fact that the spacer is non palindromic means it gives an overall orientation to the target site and dictates the way that two sites will recombine with each other (Figure 1.2). If the spacers are arranged in direct orientation on the same piece of DNA, recombination will result in excision of the intervening DNA as a circle. If the spacers of the two target sites are inverted with respect to one another, the result of recombination will be an inversion of the intervening DNA. The reverse of excision is integration of a circle carrying one target site, into a single site on another DNA molecule. If the two sites are on separate linear molecules, recombination results in reciprocal exchange of DNA between the two molecules.

Although FLP, Cre and λ integrase carry out, respectively, inversion, excision and integration as their normal *in vivo* functions, FLP and Cre can carry out any type of reaction in a reversible manner given the correct orientation of target site spacers (Senecoff and Cox, 1986, reviewed in Kilby et al., 1993). For λ integrase, the outcome of recombination is additionally determined by accessory proteins, although under suitable *in vitro* conditions, the enzyme can also catalyse inversion (Sadowski 1993). For FLP and Cre, recombination

Figure 1.2

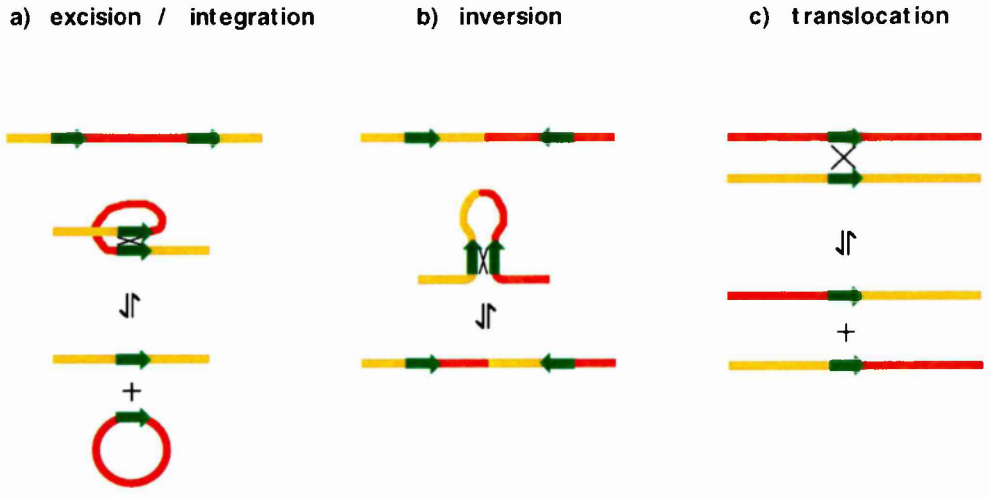


Figure 1.2 Spacer orientation determines the outcome of recombination. FRT or LoxP sites are shown as *green arrows*; the orientation of the arrow indicates the orientation of the non palindromic spacer. See text for full description of reactions.

of each type occurs with similar efficiency *in vitro* on a linear or a supercoiled substrate (Meyer-Leon et al. 1984, Abremski and Hoess 1984). However for the integration reaction of λ integrase the molecule containing attP must be supercoiled, whilst in the partner molecule containing att B, there is no requirement for supercoiling (Sadowski 1993). In *E.coli,* Cre mediated excision on plasmid substrates is greatly stimulated by negative supercoiling, and is about three times more efficient than the inversion reaction on similar substrates (Adams et al., 1992).

1.2.5 The outcome of recombination: resolvase/ invertase family

In the resolvase/ integrase family two striking difference from the integrase family are that the requirement for supercoiling is far more stringent, and that the possible outcomes of the reaction are far more limited. In this family recombination does not proceed via a Holliday junction intermediate. Instead, a double-strand break, staggered by two base pairs and with a recessed 5′ end, is made in the spacer of each target site. The ends are rearranged and rejoined in the recombination synapse (Kanaar et al. 1990, Johnson 1991, Stark et al. 1991).

Like those of the integrase family, resolvase/invertase family target sites are not perfectly palindromic, and thus they too have an overall orientation. However, in contrast to the flexibility of integrases, the resolvase enzymes can catalyse recombination only between two directly oriented target sites on the same molecule, resulting in excision of the intervening DNA. The invertases, as their name suggests, are only able to catalyse inversion of DNA between two target sites in inverted orientation.

Why is only one configuration of target sites tolerated in each case? The answer to this question came from topological studies of *in vitro* reactions. The resolvases and invertases so far characterised all require a negatively supercoiled substrate, and give recombination products of precise topology. In the case of resolvases, the products are singly catenated circles, whilst invertases give an unknotted circle. However, in all cases the

product molecules have four fewer negative supercoils than the starting substrate. This change in supercoiling represents an overall decrease in free energy, and it is this energy which is used for assembly and stabilisation of the synaptic complex during recombination. For each enzyme, only one configuration of target sites is energetically favourable. Thus the sites can be stably aligned in only one way. Interestingly, the Mu transposase, which normally catalyses strand transfer only between inverted sites, can catalyse recombination between target sites on two separate molecules provided the two substrates are sufficiently catenated and negatively supercoiled (Craigie and Mizuuchi, 1986) (for reviews, see Stark et al. 1992, Sadowski, 1993).

In addition to supercoiling, the alignment of the synaptic complex depends on accessory proteins. Resolvases bind at the two additional subsites, and invertases require the Fis protein bound at the recombination enhancer. Interestingly, several Gin invertase mutants have been identified which do not require Fis for recombination. Furthermore these mutants are able to catalyse inversions, excisions and intermolecular recombination *in vivo.* One Fis-independent mutant analysed *in vitro* (Gin K162R) was even able to carry out inversion, deletion and intermolecular recombination on linear substrates (Klippel et al. 1988). These results suggest that the limitations normally imposed on these recombinases by energy requirements can be overcome, producing enzymes which behave more like integrases.

1.2.6 Integrase family recombinases: tools for reverse genetics.

The preceding section showed that conservative site specific recombinases differ greatly in the complexity of their target sites, their requirements for accessory proteins, and the flexibility of the recombination reaction. In summary, the integrase family members Cre, FLP and R, represent the simplest systems, and also the most flexible in terms of the reactions possible. It is this combination of simplicity and flexibility which has led to the rapidly expanding use of Cre, FLP and (to a lesser extent) R recombinases for genomic engineering. In this context, the particular advantages which they

have in common are: 1) A single protein is sufficient to carry out recombination. 2) The minimal target site is quite short and simple (34 base pairs or less in all cases). This means it can be placed in many contexts in the genome without disrupting normal activity. If placed in the correct reading frame, all three sites can be read in both directions without encountering a stop codon, an essential feature if the site is to be placed in a coding sequence. 3) The target sites are long enough not to occur at random in a complex genome, and to ensure specificity of recombinase recognition. 4) Recombination is conservative. This has the advantage that its outcome is predictable at the nucleotide level. It also means that the reaction is reversible. 5) The outcome of intramolecular recombination is dictated only by the orientation of target site spacers, enabling site specific excision or inversion. 6) Intermolecular recombination is also possible, giving integration or translocation.

Incidentally, the features listed above are all shared by the Gin mutants described above. Indeed this mutant has been successfully applied to genetic manipulation in plant protoplasts (Maeser and Kahmann, 1991). However, the most widely used recombinases by far are FLP and Cre. They are also amongst the best characterised biochemically, and are the main focus of this thesis.

1.3 FLP and Cre Recombinases: Reaction Mechanism

This section will review the reaction mechanism itself in detail, using evidence from *in vitro* experiments for FLP and Cre wherever available. The general steps in FLP and Cre mediated recombination are similar (Figure 1.3). A closer look at individual steps reveals many more similarities, but also some interesting differences between the enzymes.

1.3.1 DNA binding

DNA binding is the first step in the reaction, in which the recombinase specifically recognises and binds to its target site.

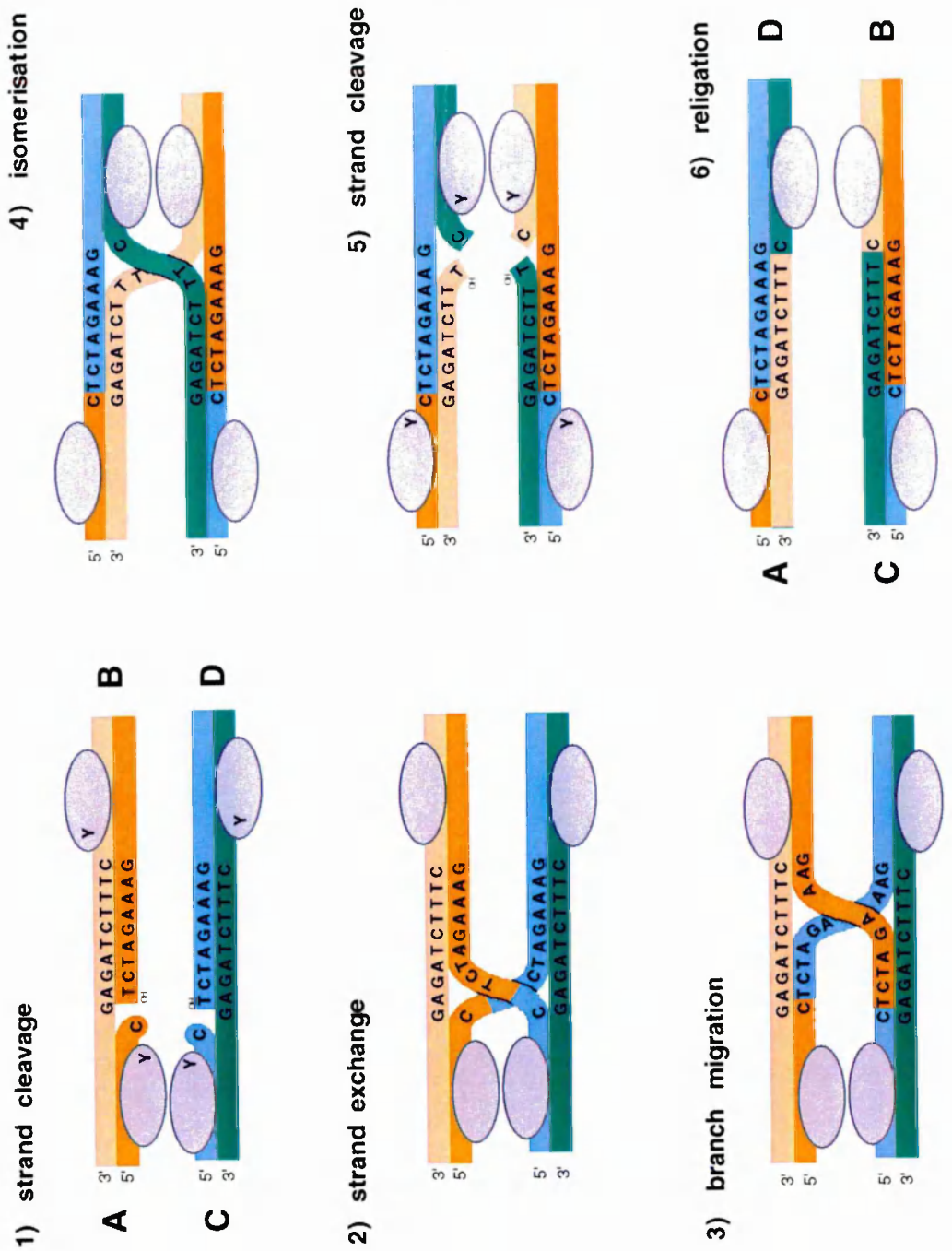
The DNA target sites of FLP and Cre are shown in Figure 1.4. The protein domains which recognise DNA are poorly defined for FLP and Cre. There is no known DNA binding motif recognisable from their aligned amino acid sequences, and no direct structural information is available. Attempts to separate the catalysis and DNA binding functions pf FLP by mutation analysis have been inconclusive (Lebreton et al., 1988, Amin and Sadowski, 1989). An extensive mutational analysis of Cre (Wierzbicki et al., 1987) showed that changes at many dispersed amino acid positions affected DNA binding *in vitro*. Since DNA binding is a prerequisite for catalysis, mutations which reduce or abolish DNA binding result in a protein with no other assayable function, and which may be impaired in folding, stability or DNA binding itself.

The DNA binding region of the very closely related integrases λ and HK022 has been examined by a domain swapping approach combined with site directed mutagenesis (the recombinses share 70% identity) (Yagil et al., 1995, Dorgai et al., 1995). 13 amino acids are necessary and sufficient to define DNA target specificity for the inverted repeat "core" sequence. One of these, at position 99 of λ, lies in the unique N- terminal region of the protein. The other 12 amino acids lie in a reasonably well conserved stretch from λ positions 279 to 329 (see Figure 1.1). FLP has not proved amenable to such an analysis, perhaps due to the lack of a very closely related partner with which to swap domains.

Using partial proteolysis, two groups have succeeded in narrowing down the specific DNA binding domain of FLP, in one case to a 21kDa peptide spanning amino acids 149 to 346. (Intact FLP is 423 amino acids long) (Pan et al., 1991 see also Figure 1.1). The other study identified a 28 kDa peptide spanning the same region (Chen et al., 1991). In the analogous experiment with Cre, a limited chymotryptic digest yielded a peptide of 25 kDa corresponding to the carboxy terminus of the protein, which was able to specifically bind the LoxP site (intact Cre is 38.5 kDa) (Hoess et al., 1990). Both the 21 kDa peptide of FLP and the 25kDa peptide of Cre map to the same region on a protein sequence alignment (see Figure 1.1). For FLP, further analysis based on footprinting and BrdU crosslinking of a 32 kDa protease cleavage fragment corresponding to amino acids

Figure 1.3

FLP reaction mechanism. The recombinase monomers are shown as grey ovals. The catalytic tyrosine is shown on panels 1 and 5 (Y). Seetext for full description of the reaction mechanism.



148 to 423, and a 13 kDa fragment corresponding to the N terminal 147 amino acids, suggest that the large C terminal fragment binds to the outer side (distal to the spacer) of the FRT half site, in a specific manner, whilst the smaller N terminal fragment binds nonspecifically to the inner part of the half site (Pan and Sadowski, 1993, Panigrahi and Sadowski, 1994). The analogous experiments have not been done for Cre, but the 25 kDa peptide was found by Fe-EDTA footprinting to protect less of the spacer region than intact Cre (Hoess et al., 1990). Taken together, these results suggest a global similarity in the modes of DNA recognition used by FLP and Cre.

Further insight into the details of DNA recognition may be inferred from recent crystal structures of the catalytic and DNA binding domains of the λ and HP1 integrases (Kwon et al., 1997, Hickman et al., 1997). Although neither structure is determined from a protein-DNA co- crystal, the position of the active site residues in relation to the rest of the structure, gives an indication of those regions of the protein likely to be in contact with DNA. For the λ integrase structure, the authors describe a basic groove, which contains all but one of the catalytic site residues as well as the amino acids previously shown to be necessary and sufficient to define DNA binding specificity (Yagil et al., 1995, Dorgai et al., 1995). This groove is therefore likely to constitute the main DNA protein interaction surface. Residues which contribute to the basic groove lie partly in the two conserved boxes I and the integrase box (Figure 1.1). A pair of basic residues at the N terminus of the catalytic fragment (figure 1.1) also line the groove, as well as a single conserved lysine residue 20 amino acids C terminal to box I. Other residues which contribute to the proposed protein DNA interface are clustered in a 50 amino acid stretch, which partially overlaps the integrase box (see Figure 1.1). This suggests that the residues making up the DNA recognition surface are dispersed throughout the protein.

For FLP and Cre, although the structural basis of DNA binding remains to be determined, the DNA target sites themselves have been relatively well characterised. As shown in Figure 1.4, the wild

type FRT site consists of an imperfect 13 base pair repeat in which the two half sites are separated by a spacer of 8 base pairs. A third direct repeat element is found 2 base pairs away from the inverted repeat. These three elements are usually referred to as a, b, and c, (marked on Figure 1.4). The third direct repeat element is not required for FLP mediated recombination *in vitro*, (Senecoff et al. 1985), although its presence appears to enhance intermolecular recombination in *E.coli* (Jayaram, 1985). The minimal Lox P site for recombination *in vitro* and *in vivo* consists of a perfect 13 base pair repeat in which the two half sites are separated by an 8 base pair non palindromic stretch (Abremski et al., 1983) (Figure 1.4).

A thorough mutation analysis of the minimal FRT site, in which each of 12 positions in the repeat element was mutated to all three other bases, has shown that mutations at 4 of the 7 nucleotides adjacent to the spacer severely reduce recombination *in vitro,* reducing recombination efficiency by up to 100 fold on substrates with the same mutation in both half sites. These mutations fall into two classes. 1) for the three nucleotides adjacent to the spacer (positions 1, 2 and 3), some changes reduce recombination severely (100 fold), whilst others have a milder effect (0 to 10 fold reduction). 2) For the G at position 7, all changes reduce recombination efficiency by at least 20 fold (Senecoff et al., 1988). This data is in good agreement with results from methylation protection and ethylation interference, which suggest that FLP makes specific contacts to bases at positions -1 to +3, at positions 5 to 7 and at positions 10 and 11 (Bruckner and Cox, 1986). No such detailed analysis has appeared for Cre, although point mutations at positions 5, 6 and 9 have no effect on recombination *in vitro* (Hoess et al. 1984). Additionally, in an approach designed to detect very low frequency (less than $1x10^{-7}$) "illegitimate" Cre mediated recombination events between Lox P and cryptic Lox sites in the yeast genome, Sauer (1992) found that recombination occurred most frequently at cryptic sites that contained the wild type sequence TATA next to the spacer, suggesting that these four positions may be important for recognition.

Figure 1.4

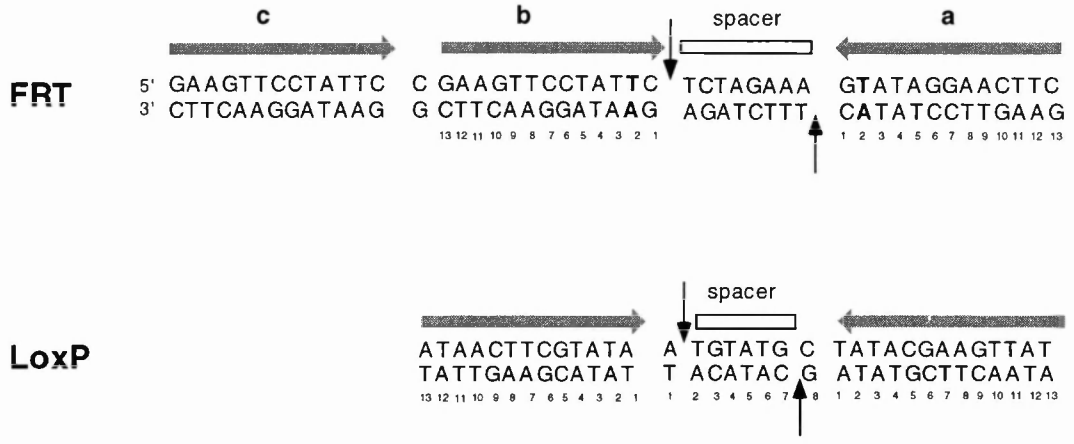


Figure 1.4 FRT and LoxP target sites.
Inverted repeat elements are shown as *horizontal arrows*. The non palindromic spacer is indicated. Cleavage points are shown by *vertical arrows*. Elements a, b,and c of the FRT site are indicated. Positions 1 to 13 of each repeat element are numbered. The single mismatch at position 2 of the FRT site is indicated by **bold type**.

In a bandshift assay, FLP forms three complexes on its wild type target site (called complexes I, II and III). DNAse I footprinting of separate gel purified complexes shows that they correspond to occupation by FLP of one, two or three elements (Andrews et al., 1987). This does not exclude the possibility that multimers of FLP bind each site. This question was addressed by comparing the three complexes with molecular weight markers, which showed that each complex represented an increment of about 50 kDa. Since the molecular weight of FLP is 48 kDa, this result strongly suggests that each 13 base pair element is bound by a FLP monomer. A later analysis confirms this result: In an adaptation of the method of Mack et al. (1992, described below), Lee and Jayaram (1995) showed that the stoichiometry of FLPY343F mutant bound to half sites was 1:1, and 2:1 in complex CII. Since these complexes comigrate with those formed by wild type FLP on the same substrate, it is reasonable to conclude that the CI FLP complex represents a monomer, and the CII a dimer.

Cre forms two complexes on its target site. In both DNAse I and chemical footprinting assays, these correspond to protection of one or both half sites of Lox P (Hoess and Abremski, 1984). The question of Cre stoichiometry was eventually addressed directly (Mack et al., 1992), by isolating each of the two complexes from a bandshift gel, and quantifying DNA and protein. This analysis showed that one molecule of Cre binds each half site.

The fact that a monomer complex exists for both FLP and Cre suggests that this is a discrete step in the binding reaction, which precedes higher order complex formation. Is there a preferred order of site occupation? For Cre, DNAse footprinting analysis with varying Cre concentrations did not show that one site was protected in preference to the other at limiting Cre concentrations (Hoess and Abremski, 1984). DNAse footprinting of FLP gave different results (Andrews et al., 1987). The monomer complex gave no DNAse protection, suggesting a weak interaction, but did present an exonuclease footprint, which showed FLP predominantly bound to element "b" (Beatty and Sadowski, 1988, see Figure 1). Complex II showed strong protection against DNAse of the b element and weaker protection of the a element (the other half of the inverted

repeat), suggesting that FLP first binds weakly to "b", then recruits a second molecule at "a", which stabilises the complex. Interestingly, the DNAse footprint of complex III showed strong protection of all three elements, suggesting that binding of a third FLP monomer at "c" stabilises the complex still further. This is in contrast to the evidence that element "c" has no effect on recombination efficiencies *in vitro* (Senecoff et al., 1985, Andrews et al., 1987).

Taken together these results show that FLP has a preferred order of assembly on its target site, and does so in a cooperative manner. The single base pair difference between sites "a" and "b" in the FRT may be sufficient to mediate this process (see Figure 1.4). It is not clear from footprinting analysis whether there is preferential occupancy and/or cooperativity for Cre. However since the two halves of the Lox P site are perfect inverted repeats, it is difficult to envisage how differential loading might occur, unless specific contacts are made to the spacer.

For both FLP and Cre, "cross core" interactions between the monomers bound to the two half sites have been demonstrated, giving support for the importance of cooperativity.

If the LoxP spacer is lengthened to 9 to 14 base pairs, Cre is no longer able to recombine homologous pairs of these molecules (Hoess et al. 1984). This suggests that cross-core interactions are very sensitive to spacer length. In an *in vitro* recombination assay, FLP tolerates modest variations in spacer length of plus or minus one base pair, provided both partner sites are homologous (Senecoff et al. 1985). FLP also forms dimeric complexes by cross core interactions in a bandshift assay on a half site substrate, as long as the "spacer" thus reconstituted does not contain more than 4 extra nucleotides (Qian et al. 1990).

Further evidence for cross core interactions comes from studies of DNA bending by FLP. Analysis by circular permutation combined with a bandshift assay showed that FLP makes three kinds of bends in its substrate DNA: 1) a "type I" bend of about 60º is caused by a monomer bound to a single half site. 2) A "type II" bend of greater

than 144° is formed by FLP binding to both halves of the inverted repeat. Because the type II bend is bigger than the sum of two type I bends, it is thought to be caused by cross core protein protein interactions. 3) A "type III" bend is made when FLP binds to elements b and c of the FRT site (Schwartz and Sadowski, 1990). Cre is reported to bend its target slightly (Hoess et al., 1984).

1.3.2 Synapsis

Once the two recombinases are bound to the target site, the next step is synapsis, in which two such bound sites come together. For both FLP and Cre, the synapse is thought to contain four recombinase monomers, each bound to a DNA half site. How do the two bound target sites find each other? The three possible mechanisms of synapsis for site specific recombinases have been extensively studied. These are random collision, tracking and slithering. These mechanisms can be distinguished by topological analysis of intramolecular recombination products, which bear the "imprint" of how the sites were brought together (Benjamin et al., 1985; Wasserman and Cozzarelli, 1985; Wasserman et al., 1985, Benjamin and Cozzarelli 1986).

In a supercoiled circular molecule, the DNA takes up a twisted conformation, in which the double helix crosses itself several times ("writhing"). If this molecule contains two recombinase target sites, it is divided into two "domains", which are wrapped around each other. If the two recombination target sites are brought together by random collision, then the two domains will remain twisted together. If the two target sites are in direct orientation, then one domain will be excised from the other, giving two catenated circles. If the reaction is an inversion, then a knot is formed. The complexity of catenanes and knots depends on the superhelical density and the distance between the recombinase target sites. If, on the other hand, target sites find each other by tracking along the DNA until they meet, so that supercoils are not trapped between them, then the product of excision would be two free circles, and the product of inversion, an unknotted circle. In the slithering model, the interwound segments move past each other

until the two sites meet across the superhelical axis, giving rise to products of distinct topology.

For FLP, the results of excision and inversion on a supercoiled circular substrate *in vitro* are entirely consistent with a random collision mechanism of synapsis. The products of excision are exclusively catenated circles, whilst inversion gives rise only to knotted products (Beatty et al. 1986).

In the same assay, Cre behaves very differently. Using purified Cre, in excisive recombination, between 59% and 73% of the products are free circles. The ratio of free to linked circles is independent of the distance between LoxP sites (from 410 to 1910 bp tested). In inversion, most of the products are unknotted (Abremski and Hoess 1985). On the basis of these results, the authors suggest that synapsis does not occur by random collision. However they also present data which argues against a slithering model, showing that on a substrate with four Lox P sites, recombination between adjacent sites is not favoured over recombination between non adjacent sites . The fact that Cre can catalyse intermolecular recombination is also evidence that random collision can occur. Data from this type of analysis does not rule out the possibility that synapsis occurs by random collision, and that products are subsequently decatenated or unknotted by Cre. Abremski et al. (1983) argue against this model, suggesting that a topoisomerase like activity would relax the DNA, which is not observed. (However, Cre was later shown to have a topoisomerase I like activity: Abremski et. al. 1986(a) and (b)). Abremski et al. (1983) also propose, but do not favour, a "nucleosome model" in which negative supercoiling is taken up by DNA winding around Cre, so that writhing is reduced (this model is disfavoured because it does not account for the production of both free and catenated products).

In a later analysis, Adams et al. (1992) used a set of differently supercoiled plasmids as substrates for Cre mediated excision *in vitro*. Their results demonstrated a linear relationship between the fraction of catenated products and the supercoil density of the substrate. The authors´ interpretation of this result is that both random collision and ordered synapsis take place *in vitro.* In the

same study, the authors show that, even in the presence of DNA gyrase inhibitors, Cre mediated excision of plasmid substrates in *E.coli* gives almost exclusively free circles. From this result, the authors conclude that synapsis occurs by an ordered process *in vivo.* However, Cre is able to catalyse intermolecular recombination though less efficiently than excision) in the same *in vivo* assay, showing that for these events, random collision must occur. Again, it is possible that the formation of unlinked products by Cre is a result not of its mode of synapsis but of a decatenation activity, which may be enhanced in the *in vivo* assay.

1.3.4 Strand cleavage

The next step in the reaction is strand cleavage, in which a single strand break is made by cleavage of a specific nucleotide at the 5´ end of the spacer, which becomes covalently attached via its 3´ phosphate to the catalytic tyrosine. Cre cleaves its substrate between the first A and T of the spacer on the top strand, and between the first G and C in the bottom strand, giving an "overlap" of 6 base pairs (Hoess and Abremski 1985), (Figure 1.4). Like λ integrase, Cre cleaves the top strand before the bottom strand (Nunes-Düby et al., 1987, Hoess et al., 1987). FLP cleaves between the the first C and the first T of the spacer in both strands (giving an 8 base pair overlap) (Senecoff et al. 1985), (Figure 1), with no preferred order of cleavage (Jayaram et al. 1988).

1.3.5 Trans Cleavage?

In Figure 1.3 panel 1, each FLP monomer is shown cleaving the nucleotide immediately adjacent to the site where it is bound. This is very probably an oversimplification of the mode of strand cleavage by FLP. In the past 5 years, much activity has focused on the question of whether FLP and other integrase family recombinases cleave their target in cis or in trans. The concept of "trans- cleavage" (Chen et al. 1992(c)) suggests that FLP uses a "shared active site", in which for one half site, the arg-his-arg triad is provided by the FLP monomer bound to that site (in cis), whilst the catalytic tyrosine is provided by a second FLP monomer, bound elsewhere in the synapse (in trans). The experiments which

gave rise to this idea were done using two different FLP mutants, one of which was mutated in one of the arg -his -arg triad residues ("triad mutant"). In the other, the catalytic tyrosine was changed to phenylalanine ("FLPY343F"), rendering it incapable of strand cleavage. The two mutants were separately bound to half sites and then mixed together, with the result that the half site bound to FLP Y343F was cleaved, showing that the catalytic tyrosine was provided in trans by the triad mutant (Chen et al. 1992(c)).

To ask whether FLP would also cleave in trans on a full site substrate, Lee et al. (1994) designed a strategy for "directed protein replacement". The assay is based on full site "suicide" substrates in which mismatches at positions 1, 2, 7 and 8 of the spacer prevent religation, thus the cleaved intermediate is trapped. In addition, one half site of the FRT contains a GC to AT change at position 7, which reduces the binding affinity of FLP (this mutation causes greater than 100 fold reduction in recombination efficiency when present in both half sites, and a 5 fold reduction when present in one half site alone (Senecoff et al. 1988). By virtue of its lower affinity for FLP, the authors show that it is possible to selectively displace prebound FLPY343F from this mutated half site by addition of an excess of wild type FLP or FLP mutant. Using this strategy, in which both the position of the cleavage deficient mutant and the point of cleavage can be monitored, the authors show that both wild type FLP and triad mutants can complement Y343F, enabling cleavage, and that this cleavage occurs preferentially in a "trans-horizontal" manner: that is, a FLP monomer cleaves the half site opposite that to which it is bound, on the same DNA molecule.

Using the analogous Cre mutants, and similar *in vitro* complementation assays, Shaikh and Sadowski (1997) showed that Cre can cleave the LoxP target in trans.

The λ integrase field was fast to ask whether this remarkable phenomenon also applied to λ int. In an experiment which preceded the "directed protein replacement" stragtegy described above, Han et al. (1993) designed full site suicide substrates, consisting of an attB half site (with low affinity for λ int) and an attP half site (with high affinity). Similarly to the FLP experiment described

above, prebound λ int Y342F (cleavage deficient) can be displaced from the low affinity sites by λ int triad mutants, giving complementation in cleavage. The complementation patterns for top strand cleavage also suggest trans horizontal cleavage.

The experiments described above were done using catalytically impaired monomers to cleave half sites, suicide substrates and substrates with reduced affinities. The question arose as to how far these reactions were relevant to normal recombination by wild type enzyme on two complete target sites. This point was addressed for λ int by Nunes Düby et al. (1994), who made use of the two closely related integrases λ and HK022. These two recombinases specifically recognise different sites, but can interact with each other to recombine hybrid sites. This system has the advantage that by designing different hybrid sites, the position of each subunit in the synapse can be predicted. Using λ integrase in combination with HK022 and hybrid sites, the authors found that cleavage was in cis and there was no evidence for a shared active site. No complementation of λ int Y342F by wild type monomer was observed. (Similar experiments with XerC and XerD came to the same conclusion (Blakely et al., 1993)).

The experiments described above show that, under certain conditions, both λ int and FLP can use a shared active site and cleave their substrate in trans. However, given a substrate more closely resembling a true target, and given the opportunity to assemble a synapse containing four catalytically competent monomers, λ int prefers to cleave in cis. The analogous experiment with FLP has not been possible since specificity mutants are not available.

These results raise the possibility that the trans cleavage observed for FLP, Cre and λ int is not the primary pathway, and is only seen when cis cleavage is blocked due to the experimental conditions used. This may occur in several ways. For example, a mutant lacking the active site tyrosine may accept a nucleophile from elsewhere, be it from another FLP monomer or from a different source. (Several authors report that FLPY343F can use small soluble "surrogate nucleophiles" for cleavage (Kimball et al. 1993, Serre at al 1993)).

Secondly, the suicide substrates used for the FLP experiments contain 4 mismatched base pairs in the spacer. Such a substrate may present less conformational constraints than wild type substrate. Finally, the mutant and wild type half sites used in these experiments differ greatly in their affinities for FLP. A synapse in which FLP monomers are bound with different affinities may not reflect the normal situation. In particular, cooperative binding of pairs of monomers is likely to be affected by the different affinities. Nunes-Duby et al. (1994) showed that cleavage of synthetic Holliday junctions requires cooperative binding of several int monomers, although this cooperation does not involve formation of a composite active site. Thus for λ int at least, efficient cleavage is dependent on cooperative binding.

The recent crystal sructures of λ and HP1 integrases have provided further insights into the phenomenon of trans cleavage. In the λ integrase structure (Kwon et al., 1997), the catalytic tyrosine is located on a flexible loop about 20 angstroms from a basic groove that contains all the other catalytic residues. This flexiblity may account for the ability of λ integrase to carry out trans cleavage under certain assay conditions. In the HP1 integrase, on the other hand, all 4 active site residues are located together in a compact globular domain (Hickman et al., 1997). Relocation of the active site tyrosine to a position suitable for trans cleavage would require considerable disruption, both of this domain and of hydrophobic interactions between monomers. This evidence strongly suggests that this integrase cleaves its substrate in cis.

1.3.6 Strand exchange

To form a Holliday junction, two cleaved strands are exchanged and religated to each other (Figure 1.3). Based on the phenotypes of "step arrest" mutants of FLP, which are inhibited at distinct steps of this process, Chen et al. (1992(c)) proposed an acid base catalysis mechanism for cleavage and ligation. The catalytic tyrosine leads a nucleophilic attack on the phosphate at the cleavage point. The hydroxyl of the tyrosine side chain may be activated by arginine 191, whilst arginine 308 may hydrogen bond

to the phosphate, making it more receptive for nucleophilic attack. Proteins with a conservative change to lysine at either of these arginine residues are able to cleave the substrate, whilst other changes abolish or greatly reduce cleavage (Parsons et al., 1990, Sadowski, 1995). Histidine 305 is not required for this step, but is essential for subsequent ligation of the cleaved strands. In this step, the 5´ hydroxyl group of the nicked strand initiates a nucleophilic attack on the phosphotyrosyl bond, and the side chain of histidine 305 may hydrogen bond to the phosphate, preparing it for nucleophilic attack. Arginine 191 has also been implicated in this step (Friesen and Sadowski 1992), and may play an analogous role to the cleavage step, by activating the 5´ hydroxyl group of the cleaved DNA strand.

1.3.7 Branch migration and the role of spacer homology

Homology in the spacer sequences of two recombining sites is not required for synapsis, but is required for recombination. (Kitts and Nash, 1987, Nunes- Düby et al., 1989, Amin et al., 1990, Hoess et al., 1986). For FLP, mutations in the spacer abolish recombination with a wild type substrate but have no effect on recombination between two identical spacer mutants, although a high AT content is preferred (Senecoff and Cox 1986, Andrews et al. 1986, Umlauf and Cox 1988, Dixon and Sadowski 1994, Schlake and Bode 1994). For Cre, point mutations at positions 6 and 7 of the spacer (see Figure 1.4) or at the three consecutive positions 6,7, and 8, allow recombination at wild type levels if both sites are mutated (Hoess et al, 1984, 1986). Conversely, a change in the AT base pair at position 5 of the spacer abolishes recombination both *in vitro* and *in vivo*, even when both recombining partners are mutated (Hoess et al., 1986).

At which point in the reaction is homology required? It has been shown for FLP and for λ integrase, that heterologies in the core region of recombining target sites cause accumulation of Holliday intermediates, showing that homology is not required for strand exchange (Meyer-Leon et al., 1990; Kitts and Nash, 1987). In addition, an att site containing a chemical cross link between

complementary nucleotides in the spacer region is able to form, but not resolve a Holliday junction (Cowart et al., 1991). These experiments are all consistent with the idea that base pairing is required for branch migration, delivering the crossover point from one end of the spacer to the other.

This idea has been challenged by experiments using synthetic Holliday junctions containing mismatches, which are nevertheless efficiently resolved in both directions by both FLP and λ integrase. (Dixon and Sadowski, 1994, Nunes-Düby et al. 1995). In both cases the authors' interpretation of the results is that Holliday junction resolution does not involve branch migration. This argument is based on the fact that such mismatched substrates are "immobilised", meaning that the branch point cannot migrate spontaneously through a mismatch on naked Holliday structures. This does not rule out the possibility that in the presence of protein, the branch point can be forced to migrate through a mismatch. The fact that a single mismatch prevents recombination but not resolution of synthetic Holliday junctions, suggests either that the requirement for spacer homology is less stringent at the resolution stage than at other stages in the reaction, or that synthetic Holliday junctions pose fewer constraints than true substrates. For Cre, the homology requirements for resolution have not been determined.

If homology is not an absolute requirement for resolution of the Holliday junction, at which stage is it required? In the case of FLP, a partial answer to this question came from an assay which separates the steps of cleavage and ligation by using nicked substrates in which the cleaved 3´phosphate is covalently attached to a lone tyrosine residue (Zhu et al., 1995). Using full site substrates, it was shown that homology in the first and second positions 5´ of the cleavage point of the FLP spacer is important for ligation of the cleaved strand to its partner. Homology at other postions is not important unless the substrates´ spacers contain single stranded regions. Similar results were obtained by Lee and Jayaram (1995). These results suggest that for FLP, base pairing at these two positions may play a role in positioning the invading strand correctly for ligation. Taking into account this 2 base pair

homology requirement for strand ligation, the "no branch migration" model outlined above, would predict for FLP that single mismatches in postions 3 to 6 would still allow recombination. However, a single change at position 3 abolishes recombination *in vitro*, (Dixon and Sadowski 1994), whilst point mutations at the other positions have not been tested. Further studies on the requirement for homology in the central four spacer nucleotides may resolve this point.

Figure 1.3 shows "isomerisation" of the Holliday junction as a distinct step. Given the complexity of the recombination synapse, this is necessarily an oversimplification. Changes in conformation of the DNA in the synapse are likely to occur comcommitantly with strand exchange and branch migration (if it occurs). However for FLP, kinetic evidence suggests that two distinct isomers of the Holliday junction may indeed exist. The argument for an isomerisation step is based on the fact that resolution of Holliday junctions but not their formation occurs much more slowly at 0°C than at 30°C. The fast formation of Holliday junctions at 0°C shows that the strand exchange and ligation reactions are not greatly affected by temperature. Therefore there must be another rate limiting step giving slow resolution at 0°C. The authors suggest that this is an isomerisation step, pointing out in addition that an increase in the AT content of the spacer increases the rate of Holliday junction resolution at 0°C (Meyer-Leon et al. 1990).

1.3.8 Resolution of the Holliday junction goes both ways.

Resolution of the Holliday junction occurs via a pair of cleavage and ligation events which are essentially the chemical reverse of the first pair (Jayaram et al. 1988), although there is evidence for FLP that R308 may play a more important role in the first pair of strand exchanges than the second (Zhu and Sadowski 1995). Both FLP and Cre resolve synthetic Holliday junctions to give an equal mix of products (Dixon and Sadowski 1993, Sadowski 1995). This means a Holliday junction formed by recombinase is equally likey to be resolved "backwards" (giving no recombination) as "forwards" (giving recombination).

There is some disagreement as to how many FLP monomers are required to resolve a Holliday junction. Bearing in mind that the "trans-horizontal cleavage/cis ligation" model would require contributions from all four FLP monomers (see Figure 1.3, steps 5 and 6), it is not surprising that this question has attracted so much attention. Using ingeniously designed synthetic Holliday junctions with mutated arms and reduced numbers of arms, three different groups showed that resolution requires two (Dixon and Sadowski, 1993 (cited in Qian and Cox, 1995), three (Qian and Cox, 1995), or four (Lee et al., 1996) FLP monomers. There is as yet no consensus, except perhaps that FLP is exquisitely sensitive to the type of substrate which it is given. Again, for Cre, no analysis of this point has appeared.

1.4 FLP and Cre: Kinetic considerations

Although the chemistry of the recombination reaction catalysed by FLP and Cre is well characterised, little kinetic analysis of the two systems has been published. This may be partly because the complexity of the reaction, and the interdependence of its steps, defies dissection into separate steps for which individual rate constants can be determined. This problem is compounded by the fact that both FLP and Cre are required in stoichiometric, rather than enzymatic amounts. For FLP which is 90% pure, the optimum molar ratio of protein monomer to half site for recombination is 10:1 (Meyer Leon et al., 1987). For Cre which is 98% pure, the optimum molar ratio of Cre monomer to half sites for recombination is 6:1 (Abremski and Hoess, 1984). These results raise the question of whether these recombinases behave as enzymes. Gates and Cox (1988) showed that FLP does behave enzymatically, albeit with a very slow substrate turnover rate of 0.12 per minute. A slow protein dissociation step was later found to be rate limiting for turnover (Waite and Cox, 1995). The fact that FLP does behave as an enzyme raises the possibility that the number of active monomers in the purified FLP preparation has been overestimated. One reason that a proportion of monomers may be inactive could be protein instability. Several authors report

problems of stability during purification of FLP. (Meyer- Leon et al., 1984, Babineau et al., 1985). Purified FLP protein tends to aggregate at NaCl concentrations of less than 0.3M, and is generally stored in 1M NaCl. Stability problems have not however been reported for Cre, which is stable during purification and can be stored in 100mM NaCl)(Abremski and Hoess, 1984). Cre is also reported to be more thermostable than FLP (Buchholz et al., 1996). Although Cre appears to be generally a more robust protein than FLP, no analysis of turnover has been published. It remains a possibility that for both FLP and Cre a certain proportion of the purified preparations is inactive.

In general, uncertainty as to the number of active monomers in a protein preparation is a major hindrance to accurate kinetic analysis. It is difficult to measure a number of events per active protein monomer when the "activity" of such a protein preparation is itself empirically determined.

1.5 Optimising the use of FLP and Cre for reverse genetics

In the previous section, the reaction mechanisms of FLP and Cre were compared, showing that there are certain differences between the two recombinases, and several unresolved questions. This section will focus on the uses of FLP and Cre as tools for *in vivo* manipulation of higher eukaryotic genomes. Much of the work published to date has been concerned with testing and optimising the FLP, R and Cre systems in cell culture and transgenic organisms, and has shown that they work efficiently in a surprisingly wide range of organisms. The huge potential of site specific recombinases for "interactive experiments" which tackle biological questions is beginning to be exploited. However the work to date also shows some limitations of these enzymes for particular purposes, and describes attempts to overcome them. This section will not give a comprehensive review of the uses of site specific recombinases, nor describe in detail the "biological" experiments which have been done. Several recent reviews cover these topics (Kilby et al. 1993, Barinaga1994, Rossant and Nagy 1995, Rajewsky

et al., 1996). Instead the following section will focus on certain aspects of optimisation, where problems remain to be solved.

1.5.1 Site specific integration

Recombinase mediated site specific integration, in which a circular plasmid bearing a single target site is integrated at a single target site previously placed on the chromosome, offers a potential alternative to homologous recombination for gene targetting. It is particularly attractive for organisms in which homologous recombination is not easily achieved (for instance in plants and flies).

Targetted chromosomal integration of a transfected plasmid has been achieved using both FLP and Cre in a variety of mammalian cell lines. In addition, site specific cointegration of two plasmids has been shown in mammalian cell culture, plant protoplast culture, and in pre-blastoderm *Drosophila* embryos (Kilby et al., 1993, Konsolaki et al., 1992). Integration, a bimolecular reaction, is kinetically less efficient than excision or inversion. This has proved true for both FLP and Cre, in various contexts. Several authors directly compare the efficiency of integration and excision, and find that integration is much less efficient than excision, both at chromosomal targets (O´Gorman et al. 1991, Logie and Stewart 1995), and on transfected plasmids (Dale and Ow 1990, Lyznik et al. 1993). The frequency of site specific chromosomal integration is particularly low, and is reported to be within the same range as random integration (Logie and Stewart 1995, O´Gorman et al., 1991).

Site specific integration is inefficient for two reasons. Firstly the integration event itself is rare, and secondly, because the reaction is reversible, the integrated DNA can be efficiently reexcised. Strategies which attempt to optimise integration, but limit reexcision have, with a few elegant exceptions, largely been based on limiting recombinase expression to a short burst, in an attempt to identify conditions in which stable integrants survive. Two alternative strategies, which differentially address both the integration and the excision efficiency, are outlined below.

By fusing FLP to the ligand binding domain of the human estrogen receptor, the recombinase activity is rendered inducible by estradiol (Logie and Stewart 1995). For FLP-EBD mediated integration of a plasmid at a single chromosomal target site in 293 cells, different protocols of estradiol administration were found to affect integration efficiencies. Of those conditions tested, the most effective was a combination of 2 hours of pre-treatment with estradiol prior to plasmid transfection, followed by 22 hours of treatment, after which estradiol was removed by washing. 22 hours was found to be more effective than 28 hours. The authors suggest that the pretreatment may serve to load the chromosomal target site with recombinase, so that it is ready to recieve the incoming plasmid, whilst the shorter post transfection treatment may reduce re- excision. Post transfection treatments of less than 22 hours were not tested.

Using Cre in tobacco protoplasts, Albert et al. (1995) describe a strategy for favouring integration over excision based on mutated LoxP target sites. Mutations are made in only one half site of each loxP site. The sites on the two DNA molecules are oriented so that recombination between them will give one Lox P site which has mutations in both halves (double mutant site), and the other which has no mutations (wild type site). If the double mutant site thus formed has a sufficiently low affinity for the recombinase, then recombination between this and the wild type site would be reduced. In this way, the integration reaction might be favoured over re-excision. The authors show three mutant LoxP sites for which this is indeed the case. In transient assays in tobacco protoplasts, the efficiency of plasmid cointegration can be increased to 7.5 times that of excision under certain conditions. The efficency of site specific integration at a chromosomal Lox target was also improved by this strategy, albeit only to a level approaching that of random integration. The effect of recombinase concentration on the plasmid co integration and excision reactions was tested by transient transfection in protoplasts. Interestingly, the integration efficiency increased with recombinase concentration, but was saturated at a certain point whereas the excision efficiency increased linearly over the Cre concentrations

tested, with the consequence that, even with the most effective mutant sites, the overall relative efficiency of integration over excision was decreased at higher Cre concentrations.

A similar concept has been demonstrated with FLP in mammalian cell lines (Schlake and Bode, 1994, Siebler and Bode, 1997). In this approach, FLP is used to induce a double reciprocal exchange event between FRT flanked cassettes at a genomic locus and on a transfected plasmid. By using appropriate combinations of mutant and wild type sites, the authors show that desired events can be favoured.

The above experiments underline the importance of finding conditions of recombinase expression or activation that optimise integration whilst not allowing too much excision. This window will depend to a large extent on the particular context of the experiment, but also on inherent characteristics of the recombinase itself, including its affinity for DNA, the mode of synapsis, the stability of the synapse and the stability of the enzyme, parameters that have mostly not been determined for FLP or Cre.

At saturating recombinase concentration, one may assume that the rate limiting factor for integration is the physical distance between the two target sites which must find each other. It follows from this that the distance between two target sites on the same chromosome could also affect the efficiency of recombination between them. Furthermore it may be possible to influence the relative efficiency of integration versus excision, by adjusting the size of the inserted DNA: the size of the plasmid to be integrated should have no effect on the rate of synapsis, but the rate of synapsis between linearly arranged target sites on a chromosome should decrease with increasing distance. There will also be a minimum optimum distance. For Cre it has been shown *in vitro* that excisive recombination is inhibited when the distance between Lox P sites is below 116 base pairs (Hoess et al 1985). No systematic evaluation of the effect of longer distance between target sites has been published for either FLP or Cre, although intramolecular recombination over very long distances (ranging from 90kb to 16.5cM) has been reported for both enzymes in various organisms,

and where frequencies are measured, they are usually extremely low despite positive selection (Osborne et al. 1995, Ramirez- Solis et al. 1995, Golic 1994). Golic and Golic (1996) observed that the frequency of recombination between FRTs on the same chromosome in *Drosophila* decreases as the distance that separates them increases.

In summary, site specific integration is possible if combined with effective positive selection procedures, and its efficiency can be improved by the design of target sites and/ or the control of recombinase expression or activity.

1.5.2 Control and recombinase efficiency

The quest for methods of switching recombinase activity on or off is not only driven by the need to prevent re-excision of integrated contructs. In transgenic organisms, control of site specific recombinase activity is desirable as a means to introduce mutations in a given tissue, or at a given time, which would otherwise be lethal. Temporal control also offers a means to mark single cells for lineage analysis. Control of a recombinase comprises two components: repression and induction, whose relative importance varies depending on the application. In this respect, the applications of site specific recombinases can be divided into two broad classes: 1) Experiments such as tissue specific or inducible gene knockouts in mice, in which recombination results in a loss of function phenotype, requiring quantitative recombination of a cell type, and in which some background recombination before induction is acceptable (Gu et al. 1994, Kühn et al. 1995). 2) Applications in which recombination results in a gain of function phenotype, in which partial recombination or even a single event is sufficient, and in which complete absence of recombination before induction is necessary. Examples of this type of experiment are cell lineage studies (Golic and Lindquist 1989), and recombination induced tumorigenesis (Pichel et al., 1993). Several different methods of controlling recombinase activity have been reported, with varying results in

terms of the extent of repression, and the efficiency of excisive recombination upon induction. Some examples are described below.

In cell culture, supplying recombinase by means of transient transfection, or lipofection of purified protein (Baubonis and Sauer, 1993) ensures that no recombination takes place before delivery. An increase in recombinase concentration gives high levels of excision, sometimes to 100%. In an experiment that could be seen as the whole organism equivalent of transient transfection, recombination in mice transgenic for lox P target sites was induced by infection with adenovirus expressing Cre (Wang et al., 1996). Excision measured by southern blot analysis reached 100% in some tissues (e.g., liver) and was partial in some other tissues.

Strategies in which the recombinase coding gene is introduced stably into the genome and is transcriptionally controlled have been widely used in transgenic organisms. In mice transgenic for Cre under the control of a thymus specific promoter, Cre expression was limited to thymocytes, where it efficiently excised a reporter gene in 87% to 99% of T-cells. No background recombination was detected in other tissues (Orban et al., 1992). Interestingly, in the "real" experiment, in which the same Cre expression strategy was used to knock out a segment of the DNA Polymerase ß gene in thymocytes, recombination occured in only 40% of the T-cell population. (No recombination was detected in other tissues) (Gu et al., 1994). These results suggest that the efficiency of recombination depends not only on the expression strategy used, but also on other factors, such as the target locus itself, where local chromatin structure may affect the accessibility of recombinase target sites.

Inducible promoters have been extensively used for the temporal control of recombinase expression. One interesting example in mice involves the LoxP containing allele of the DNA polymeraseß gene described above ("polß^{flox}"). Kühn et al. (1995) used an interferon responsive promoter to drive Cre expression in transgenic mice, resulting in excision of polß^{flox}. After administration of interferon, excision efficiencies were checked by Southern blot analysis of different tissues, showing that recombination occurred

in 94% to 100% of spleen and liver cells, but was reduced in other tissues. Without interferon treatment, excision levels in the spleen and liver were 10% and 2% respectively.

In plants and flies, inducible transcription of FLP has been achieved using heat shock promoters, which have proved a reliable means of providing conditional gene expression. In flies, the *Drosophila* hsp70 promoter has been used to drive FLP expression (Golic and Lindquist 1989, Golic., 1991). By using strategies in which recombination gives rise to expression of a visible marker gene, single cells and their descendents can be marked at selected developmental stages by heat shock induced transient expression of FLP. This type of experiment requires that FLP is not active before heat shock, that it does not remain active for long periods afterwards, and that efficiency of recombination is low. The reported thermolability of FLP may account for its ablilty to fulfill these requirements (Buchholz et al., 1996). In the first reported use of Cre in flies, Siegal and Hartl (1996) expressed Cre under the dual Hsp70/mos1 promoter, showing virtually 100% excision of a white reporter gene in somatic and germline cells. This result is consistent with the thermostability reported for Cre (Buchholz et al., 1996).

Tissue specific expression of FLP has been achieved in flies by use of the β2 tubulin promoter (Kosman and Small, 1997), and the GAL4 UAS system (Lin et al., 1995).

In maize protoplasts (Lyznik et al. 1995) and *Arabidopsis* seedlings (Kilby et al. 1995), FLP has been expressed using Soybean heatshock promoters. No background recombination is detectable at 20°C. Authors report up to 25% excision upon heat shock of 2 to 4 hours at 37 to 42°C. Again, it is possible that the thermolability of FLP contributes to the low recombination efficiency seen, and that this could be optimised for applications where more recombination is required.

An alternative to transcriptional control is hormone induction of recombinase activity by virtue of its fusion to the ligand binding domain of a nuclear hormone receptor. The FLP-estrogen binding domain fusion described above (Logie and Stewart 1995) spawned a

spate of similar fusion proteins, in which Cre was fused to a ligand binding domain, and thus rendered inducible by the appropriate ligand (Metzger et al. 1995, Kellendonk et al. 1996, Zhang et al. 1996). The fusion proteins have to date been characterised in mammalian cell culture, (including ES cells in some cases), in yeast (Nichols et al., 1997), and in transgenic mice, where recombination activity was rendered dependent upon administration of tamoxifen (Feil et al., 1996, Schwenk, F. Kühn, R., Angrand, P.-O., Rajewsky, K., and Stewart, A.F., submitted).

In summary, the experiments reviewed here suggest that in order to design an effective strategy for a particular recombinase based *in vivo* experiment, it is important to optimally select 1) the induction strategy, 2) the recombinase, and 3) the target locus.

Regarding selection of the recombinase itself, little has been done to characterise the properties of FLP, Cre and R which may be relevant to their applied use *in vivo.* For instance, these may include affinity for DNA, the stability of the synapse, tolerance of distance between target sites, and stability of the enzyme itself. Regarding this last point, the activity of FLP and Cre at different temperatures has been tested (Buchholz et al. 1996) showing that FLP has reduced activity at 37°C and above *in vitro,* in *E.coli,* and in mammalian cell culture. The authors suggest that this result may make FLP less suitable than Cre for use in transgenic mice (whose body temperature is 39°C). However FLP has since been shown to function efficiently in transgenic mice (Dymecki 1996). A FLP transgene was placed under the control of the human ß actin gene promoter, and gave efficient recombination of a LacZ reporter gene in a wide range of tissues, with efficiencies of 30 to 78%. These levels are comparable to those reported for Cre (Orban et al., 1992, Gu at al., 1994). It is probable that by expressing high enough levels of FLP the effect of temperature on its activity is overcome.

Regarding the selection of the target locus, it is likely that chromatin structure will have an effect on recombination efficiencies. The difference in Cre mediated recombination efficiencies in T- cells of transgenic mice at two different loci (Orban et al., 1992, Gu et al., 1994), suggests that chromatin

structure may play a role. It has been suggested that due to its prokaryotic origins, Cre may be less able to recombine its target site in a chromatin context than FLP, which comes from a eukaryote. The fact that both recombinases can function effectively in higher eukaryotes does not support this idea, but a comparative evaluation of the effects of chromatin on the activity of FLP and Cre has not appeared.

1.5.3 The future: the right recombinase(s) for the right job.

The uses of site specific recombinases are expanding, both in range and in subtlety. An understanding of the inherent properties of different recombinases would enable them to be used optimally for different strategies. Eventually a detailed understanding of the parameters which affect different aspects of the recombination reaction may enable mutants to be designed for specific purposes.

The choice of the right recombinase for a particular application is limited at present to three enzymes. The identification of new recombinases, which may have different properties, would increase the choice available. In addition, experiments in which more than one recombinase are used are becoming feasible. For example, a first recombinase could be used in cultured ES cells to delete a selectable marker from a targetted gene, and a second in transgenic mice for a conditional or inducible knockout. Potentially, experiments involving multiple independent recombination events are possible. In this context, the identification and characterisation of novel recombinases is desirable.

1.6 The aims of this thesis

The overall aim of this thesis is to investigate and compare the biochemical properties of site specific recombinases, with a view to optimising their use in genomic manipulation strategies.

Although the reaction mechanism of many integrase family site specific recombinases is well characterised, little attention has been paid to the inherent properties of these recombinases which may be relevant to their applied use for *in vivo* genomic manipulation. Furthermore, although FLP and Cre recombinases have both been used extensively for *in vivo* experiments, little comparative analysis of the two recombinases has been published. Finally, at present the choice of recombinases available for *in vivo* experiments is limited to three enzymes: FLP, Cre and R.

The first aim of this thesis is to characterise a novel integrase family recombinase, the Kw recombinase from the yeast *Kluyveromyces waltii*, and to investigate its suitability for use in *in vivo* experiments. Using *in vitro* assays, I aim to compare the properties of Kw recombinase with those of FLP and Cre.

Secondly, I aim to undertake a thorough comparative analysis of FLP and Cre. Using *in vitro* assays and purified proteins, precise quantitative comparisons can be made, thereby identifying differences in the inherent properties of the two recombinases. The properties which are the most interesting from the perspective of reverse genetics are DNA binding affinity, and the protein protein interactions which dictate the formation and stability of the synaptic complex. Using a combination of kinetic and mathematical approaches, I aim to characterise and quantify these processes for FLP and Cre, and to see how they differ for the two proteins.

Another characteristic which is directly relevant to the *in vivo* application of site specific recombinases, is the temperature sensitivity of the enzyme, because the growth temperature of the host organism dictates the temperature at which the experiment is carried out. For this reason, I aim to compare the thermostabilities

of FLP, Cre, Kw and two FLP mutants, and to investigate reasons for the behaviour observed in each case.

Finally, there has been no evaluation of the effect of distance between target sites on recombinase activity. Therefore, in my last project I aim to determine the minimum and maximum optimum distance between FRT sites for FLP mediated excision *in vitro*, and to determine parameters which can affect these limits.

Throughout this thesis, I aim to discuss the results both from the point of view of recombinase action, and in the context of their potential relevance for the optimisation of site specific recombinase technology.

CHAPTER 2
MATERIALS & METHODS

2.1 Materials

2.1.1 Chemicals

Bacterial media were bought from Gibco-BRL, USA. Agarose was purchased from Sigma, Germany. 30% acrylamide: bis- acrylamide solution was bought ready prepared from National Diagnostics, USA. Sodium dodecyl sulphate (SDS) was from Biorad, Germany. Poly dIdC, deoxyribonucleoside triphosphates (dNTPs) and ribonucleoside triphosphates (NTPs) were purchased from Boehringer Mannheim, Germany. α-thiol dNTPs were from Pharmacia, Sweden. Unless otherwise specified, all other organic and inorganic chemicals were obtained from Biorad, Boehringer Mannheim, Fluka, Merck, Serva or Sigma, Germany.

2.1.2 Radioactive isotopes

^{35}S-α-dATP (5000Ci/mmol), ^{32}P-α-dCTP, ^{32}P-α-dGTP, ^{32}P-α-dATP and ^{32}P-α-dTTP (all 3000Ci/mmol), ^{32}P-α-UTP (800Ci/mmol) and ^{35}S-Methionine (1000Ci/mmol) were purchased from Amersham, UK.

2.1.3 Enzymes

Restriction enzymes were bought from New England Biolabs, USA, Fermantas, Lithuania, or Boehringer Mannheim, Germany. *E.coli* DNA polymerase I (Klenow fragment) and T4 DNA ligase were from New England Biolabs, USA. Calf intestinal phosphatase (CIP) was bought from Boehringer Mannheim, Germany. T4 polynucleotide kinase, T3 and T7 RNA polymerases were obtained from Stratagene, Germany. RQ1 DNAse I and recombinant human RNAsin were from Promega, USA. RNAse A, Proteinase K and Bovine serum Albumin (BSA) were bought from Sigma, Germany. Acetylated BSA was from New England Biolabs, USA. Amplitaq^{TM} DNA polymerase was bought from Perkin Elmer Cetus, USA.

The purified FLP and Cre preparations used in this study are >95% pure as judged by SDS-PAGE. Cre protein was purified by Frank Buchholz as described (Abremski and Hoess, 1984). FLP protein was a generous gift from Michael Cox, and was purified as described (Meyer-Leon et al, 1987). The activity of both preparations is assumed to be 100%.

2.1.4 Synthetic oligonucleotides

All the oligonucleotides used in this work were synthesised by the oligo service at EMBL.

2.1.5 Other materials

Nuclease treated rabbit reticulocyte lysate, complete amino acid mix and amino acid mix minus methionine were obtained from Promega, USA. Kodak XAR5 X-ray film was from Eastman Kodak, USA.

2.2 Bacterial techniques

2.2.1 Bacterial strains

The *E.coli* strains used for the propagation of double stranded plasmids were:
HB101 (genotype: *supE44 ara14 galK2 lacY1 Δ(gpt-proA)62 rpsL20 (Str^r) xyl-5 mtl-1 recA13 Δ(mcrC-mrr)* HsdS$^-$ (r^-m^-))
and XL1-blue (genotype: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* (F′ *proAB lacI^{q}ZΔM15* Tn10 (Tet^r))^c

2.2.2 Maintenance and media

Bacterial cultures were grown at 37°C in L-broth, cells were grown on L-broth agar plates as described (Sambrook et al, 1989) except that ampicillin was added at 100μ g/ml.

2.2.3 Preparation and transformation of competent cells

HB101 or XL-1 blue were made competent for transformation as follows. 1ml of an overnight culture was diluted into 100ml L-broth without antibiotics and grown at 37°C until an OD_{600} of 0.6 was reached. All subsequent steps were done at 4°C in the cold room. After chilling in an ice water bath for 10 minutes, the bacterial culture was transferred to pre-chilled 50ml Falcon tubes and centrifuged for 10 minutes at 4°C at 3,000 rpm in a Sorvall centrifuge. The pelleted cells were resuspended in 25ml of pre-chilled 0.1M $MgCl_2$ by gently pipetting up and down. After 20 minutes on ice, the cells were harvested as described above, gently resuspended in 25ml of pre- chilled 0.1M $CaCl_2$ and incubated on ice for a further 30 minutes. Cells were again harvested as above and resuspended in 4ml of pre-chilled 0.1M $CaCl_2$ containing 14% glycerol. 100 µl aliquots were frozen in liquid nitriogen and stored at -70°C. Competence of cells was determined by transforming with a range of known quantities of supercoiled plasmid.

For transformation of competent cells, aliquots were thawed on ice and mixed with ligation mix, not exceeding 1/10 of the volume of competent cells. After incubation on ice for 30 minutes, cells were heat shocked for 2 minutes at 42°C, then diluted ten fold into L-broth without antibiotics and incubated for 45 minutes at 37°C in a shaking block. Cells were harvested by centrifugation for 30 seconds in an Eppendorf centrifuge, resuspended in 50µl L- broth, and spread on L-agar plates pre- warmed at 37°C and containing appropriate antibiotics. Plates were incubated at 37°C overnight.

2.3 Recombinant DNA techniques

2.3.1 Restriction enzyme digestions

Restriction enzyme digestion of double stranded DNA was carried out according to the instructions of the supplier, using the buffer provided.

2.3.2 Other enzymes

CIP (calf intestine phosphatase) was used where appropriate to remove 5´ phosphates from vector DNA, thus reducing background caused by recircularisation of vector. EDTA was added to the restriction digest to 20mM, 1/100 volume of CIP was added, followed by incubation at 37°C for 10 minutes. CIP was inactivated by incubation at 70°C for 15 minutes, followed by phenol extraction and ethanol precipitation. Blunt ends were generated from 5´overhangs by Klenow polymerase. 1/10 volume of dNTP mix (2.5mM of each dNTP) and 0.5µl of Klenow polymerase were added to the restriction digest, and incubated for 10 minutes at 23°C, then for 10 minutes at 65°C.

2.3.3 Gel purification of DNA fragments from agarose gels

DNA fragments to be ligated were purified by electrophoresis on agarose minigels in 1XTBE, the desired band was visualised under UV light at 312 nm and excised. DNA was extracted using a QIAEX II kit (Qiagen, Germany) according to supplier´s instrustions. When purifying fragments for use as recombination substrates (section 2.8.2) or as templates for the synthesis of DNA probes for southern hybridisation (section 2.5.3), 1X TAE gels were used (section 2.4.1).

2.3.4 Preparation of oligonucleotides for subcloning

Complementary pairs of oligos were phosphorylated and annealed in 50µl volume by mixing 50 pmoles of each oligo with 1µl polynucleotide kinase in kinase buffer (70mM Tris-HCl pH7.6, 10mM $MgCl_2$, 5mM DTT, 1mM ATP. Reactions were incubated at 37°C for 1 hour, EDTA was added to 10mM, then the sample was heated to 70°C for 15 minutes and allowed to cool slowly to room temperature. This preparation was used directly for ligations.

2.3.5 Ligations

Ligations were carried out by mixing vector and insert DNA (with insert in at least 3 fold molar excess over vector), in a volume of 10 to 20µl ligation buffer (66mM Tris-HCl pH 7.5, 6.6mM $MgCl_2$, 1mM DTT, 1mM ATP) with 1unit of T4 DNA ligase. Ligations were incubated at 23°C for 2 to 24 hours.

2.3.6 Mini preparation of plasmid DNA

Small scale DNA preparations for analytical purposes were made from 1ml overnight cultures by the alkaline lysis method. Cells were pelleted by 30 seconds contrifugation in an Eppendorf centrifuge, and all medium was removed. The pellet was resuspended in 100µl solution I (50mM glucose, 25mM Tris-HCl pH8.0, 10mM EDTA) and mixed by vortexing. 200µl solution II was added (0.2N NaOH, 0.1% SDS) and the tubes were placed on ice. 150µl solution III was added (3M KAc, 11.5% acetic acid) followed by incubation on ice for 3 to 5 minutes. Cell lysates were cleared by centrifugation for 5 minutes at 4°C at 14,000 rpm in an Eppendorf centrifuge. Plasmid DNA was recovered from the supernatant by phenol extraction and ethanol precipitiation. Pellets were washed in 70% ethanol and resuspended in appropriate restriction enzyme buffer. For analysis of digests, RNAse A was added to the gel loading buffer to a final concentration of 0.1mg/ml.

2.3.7 Maxi preparation of plasmid DNA

Large scale preparations of plasmid DNA were made from 500ml overnight cultures using a plasmid maxi kit (Qiagen, Germany). Bacterial cells were pelleted and resuspended in 10ml of buffer P1 (50mM Tris-HCl pH8, 10mM EDTA, 100µg/ml RNAse A), then lysed by mixing with 10ml buffer P2 (0.2N NaOH, 1% SDS) and incubated at 23°C for 5 minutes. 10ml of buffer P3 were added (2.55M KAc, pH4.8) followed by 20 minutes incubation on ice. The lysate was then cetrifuged for 30 minutes at 4°C at 8,000 rpm in a Sorvall SS34 rotor, and the supernatant loaded onto a Qiagen 500 column, previously equilibrated with 750mM NaCl, 50mM MOPS pH 7.0, 15% ethanol, 0.15% Triton X-100, and allowed to flow through. The

column was then washed twice with 30ml of 1M NaCl, 50mM MOPS pH7.0. DNA was eluted with 15ml of 1.25M NaCl, 10mM Tris-HCl pH8.0, 1mM EDTA, and precipitated with 11.5ml of isopropanol. The sample was centrifuged for 30 minutes at 8,000 rpm and the DNA pellet washed with 70% ethanol, air dried and resuspended in TE (10mM Tris-HCl pH 8.0, 0.1mM EDTA). The OD_{260} of the plasmid solution was determined and the preparation was diluted to 1mg/ml.

2.3.8 Sequencing of plasmid DNA

DNA was sequenced by the dideoxy method using the Sequenase™ kit (USB). 2-4 µg of plasmid DNA was denatured by dilution to 18µl with distilled water and addition of 2µl 2M NaOH, and incubated at room temperature for 5 minutes. DNA was precipitated and resuspended with 0.5 pmol sequencing primer in sequenase buffer (USB) in a volume of 10µl The primer was annealed by heating to 68°C for 2 minutes and cooling slowly to room temperature. Meanwhile, eppendorf tubes were prepared for termination, containing 2.5µl of each of the dideoxynucleotide termination mixtures (G, A, T and C) provided in the USB kit. Termination tubes were pre warmed at 37°C. The labelling mix provided in the USB kit was diluted 1/5, and the Sequenase™ enzyme was diluted 1/8 in enzyme dilution buffer. The following components were then added to the annealed DNA mixture: 1µl DTT (USB), 2µl diluted labelling mix, 1µl ^{35}S-α-dATP and 2µl diluted Sequenase™. Samples were incubated at 23°C for 2 to 5 minutes, then 3.5µl was transferred to each termination tube (G, A, T, C), and incubated for a further 5 minutes at 37°C. 4µl Stop solution (USB) was added, and samples were heated at 80°C for 2 minutes before loading on a 6% denaturung polyacrylamide gel.

2.3.9 Polymerase chain reaction (PCR)

PCR reactions were carried out in 100µl volume containing 50mM KCl, 10mM Tris-HCl pH 6.8, 0.5mM MgCl2, 0.2mM dNTPs, 10ng plasmid template, 50pmol of each primer and 2.5U Amplitaq™ DNA

polymerase. Samples were overlaid with 100µl mineral oil to prevent evaporation. DNA amplification was carried out in a thermal cycler with 30 cycles of 1 minute denaturation at 95°C, 2 minutes at appropriate annealing temperature determined for individual primers (37 to 60°C), 2 minutes extension at 72°C. The last cycle was followed by a 5 minute incubation at 72°C to fill in ends. To prepare PCR products for subcloning, the aqueous layer was removed and precipitated in the presence of 1µl glycogen (Boehringer Mannheim), then digested with appropriate restriction enzymes and the desired fragment gel purified. All PCR generated constructs were sequenced.

2.3.10 Site directed mutagenesis

Unless otherwise stated, point mutations were made by overlap extension PCR, (Horton et al, 1989), which is based on the following principle: the DNA sequence of interest is amplified by PCR in two sections whose ends overlap by 15 base pairs at the point where the mutation is to be introduced. The mutation is introduced into two primers complementary to both strands of this 15 base pair region, and thus complementary to each other. Amplification of the two sections, using each of these two primers in combination with a primer for the other end of the section to be amplified, in separate PCR reactions generates molecules whose ends are complementary to each other. The fragments are gel purified and mixed in a second PCR reaction containing only the primers for the extreme 5´and 3´ ends of the full DNA sequence. The two fragments with complementary ends thus anneal to each other and act as primers for the first round of amplification of the entire sequence, containing the introduced mutation.

2.4 Electrophoresis

2.4.1 Agarose gel electrophoresis

Agarose gel electrophoresis, used for analysis of RNA and analysis and purification of DNA fragments of 100 base pairs and above, was

essentially as described by (Sambrook 1989). The required weight of agarose for 0.6 to 2% (w/v) gels was dissolved in 1X TBE (90mM Tris-HCl pH 8.3, 90mM boric acid, 1mM EDTA) or 1X TAE (90mM Tris-HCl pH 8.3, 90mM acetic acid, 1mM EDTA) by boiling for several minutes. The solution was cooled and ethidium bromide was added to a concentration of 0.05µg/ml just before pouring. Before loading samples, 1/10th volume of agarose gel sample buffer was added (25% sucrose, 1mM EDTA pH8.0, 0.1% (w/v) bromophenol blue, 0.1% (w/v) xylene cyanol). The gels were run submerged in 1XTBE or 1X TAE as appropriate at 15-20 volts/cm (TBE minigels) or 2.5-5 volts/cm (20X20cm gels and TAE minigels).

2.4.2 Non- denaturing polyacrylamide gels

Native polyacrylamide gel electrophoresis was used for mobility shift analysis and also for purification of radioactively labelled double stranded DNA fragments of 70 base pairs and below.

Gels were poured between two glass plates (20X34 cm), with 1.5mm spacers, clamped together and sealed with rubber tubing. The gel solution was made in 100ml in 1X TBE buffer with the following concentrations of 30% acrylamide:bisacrylamide 29:1:16.7ml for 5% gel used in bandshift analysis; 26.7ml for 8% gel used for purification of DNA. 600µl 10% ammonium persulphate and 100µl TEMED were added, the gel was poured and allowed to set for 45 minutes. The gel was pre run in 1X TBE for 1 hour at constant voltage (250V) in the cold room before samples were loaded.

2.4.3 DNA sequencing gels

Two glass plates (20X40cm) cleaned with ethanol and treated with dimethyldichlorosilane solution were taped together with 0.4mm spacers. The gel solution was made in 50ml 1XTBE and contained 21g urea, and10ml 30% acrylamide: bisacrylamide 20:1 (giving a 6% gel). 300µl 10% ammonium persulphate and 50µl TEMED were added to polymerise, the gel was poured and allowed to set for 45 minutes. The gel was prerun in 1X TBE for 1 hour at room

temperature at constant power (40W). The samples were then loaded and electrophoresed at 40W for the appropriate time (1.5 to 4 hours). After electrophoresis the gel was fixed in 10% acetic acid, 10% methanol for 30 minutes before drying down on Whatmann 3MM paper and exposure to Kodak XAR5 film at room temperature.

2.4.4 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used for the analysis of proteins and was carried out as described by Sambrook et al (1989). Gels were poured between two glass plates (10X12cm) separated by 1mm spacers clamped together and sealed with rubber tubing. The separating gel contained 8 to 12% acrylamide: bisacrylamide 29:1, 375mM Tris-HCl pH8.8, 0.1% SDS in a volume of 10ml, and was polymerised using 90µl ammonium persulphate and 10µl TEMED. The separating gel was ovelraid with butanol and allowed to set for 20 minutes, after which the butanol was removed and the stacking gel was poured, containing 4% acrylamide: bisacrylamide 29:1, 125mM Tris-HCl pH 6.8, 0.1% SDS in a volume of 10ml, polymerised as above. Before loading, samples were mixed with 1/5 volume 6x sample buffer (0.25mM Tris-HCl pH 6.8, 36% (w/v) glycerol, 10% (w/v) SDS, 0.12% (w/v) bromophenol blue, 600mM DTT), and heated to 80°C for 2 minutes to denature proteins. Gels were run at constant voltage (200V) in running buffer (25mM Tris-HCl, 200mM glycine, 0.1% SDS).

After electrophoresis, gels containing ^{35}S-methionine labelled proteins were dried down and autoradiographed. Gels containing pure proteins were stained for 30 minutes in 50% methanol, 0.05% (w/v) Coomassie brilliant blue R250 (Serva), 10% acetic acid), and destained in 40% methanol, 10% acetic acid for several hours.

2.5 Southern blotting

2.5.1 DNA transfer

After electrophoresis, DNA gels were incubated with gentle shaking in 0.25M HCl for 9 minutes. This results in partial depurination of

DNA. The gel was then transferred to 0.4M NaOH for three 15 minute washes. This causes denaturation of DNA and conversion of depurinated bases into single strand breaks, thus facilitating transfer. The gel was then incubated in 20x SSC (3M NaCl, 0.3M $C_6H_5Na_3O_7.2H_2O$) for 20 minutes prior to blotting onto a Qiabrane nylon membrane (Qiagen). The gel was inverted and DNA transfer was mediated by capillary action in 20x SSC overnight. The membrane was then rinsed in 25mM NaH_2PO_4 pH7.2, 1mM EDTA, allowed to dry, and exposed to UV radiation at 254nm for 30 seconds, to crosslink the DNA to the membrane. The filter was then rolled and placed in a glass tube (30cm x 5cm) and prehybridised in 8 to 10 ml Church and Gilbert buffer (7% SDS, 250mM NaH_2PO_4, 1% BSA, 1mM EDTA) for 30 minutes at 72°C in a hybridisation oven fitted with a rotating wheel.

2.5.2 RNA probe

RNA probes were made from DNA templates carrying a T3 or T7 promoter. 1µg of template DNA linearised with an appropriate restriction enzyme was incubated in 15µl volume with 3µl 5x transcription buffer (Stratagene), 2µl 3.3mM each ATP, GTP and CTP, 6µl ^{32}P-α-UTP (800Ci/mmol), 1µl RNAsin and 20 units T3 or T7 polymerase. The reaction was incubated at 37°C for 15 minutes, followed by addition of 40 units of RQ1 DNAse 1 and further incubation at 37°C for 15 minutes. Incorporation of radioactive label was monitored by spotting 0.2µl of the reaction onto 0.1mm cellulose polyethyleneimine thin layer chromatography (TLC) paper before and after addition of the RNA polymerase enzyme. Unincorporated and incorporated UTP were resolved by chromatography in 0.75M KPO_4 pH3.5, and the chromatograph dried and autoradiographed for 1 minute. After DNAse treatment, 40µl of Church and Gilbert buffer were added to the reaction, and the probe was added directly to the prehybridised filter.

2.5.3 DNA probe

Where no suitable plasmid was available from which RNA could be sythesised, DNA probes were made by random priming using the "prime- a- gene" kit from Promega as follows: A DNA restriction fragment of 500 to 800 base pairs containing the sequence of interest was purified twice by agarose gel electrophoresis in a 1XTAE gel and resupended at 10ng/μl. 25ng of DNA in 9μl volume were denatured by heating to 95°C for 5 minutes then immediately transferred to ice. 5μl of 32P-α-dCTP (3000Ci/mmol), 1μl Klenow polymerase, and 5μl of Random primer mix supplied in the kit, containing reaction buffer, DNA primers, dATP, dGTP and dTTP, were added and the reaction was incubated at 37°C for 30 minutes. Incorporation of radioactive dCTP was monitored by TLC as described above. Unincorporated dNTPs were removed by precipitation and the probe was added to the prehybridised filter.

Both RNA and DNA probes were hybridised for 16 hours at 72°C, after which the filter was rinsed in wash buffer (20mM NaH_2PO_4,1% SDS, 1mM EDTA) at room temperature, then incubated for 1 hour in the glass tube in wash buffer at 72°C in the hybridisation oven, prior to autoradiography or PhosphorImager analysis.

2.6 *In vitro* transcription and translation

2.6.1 *In vitro* transcription

pBluescribe and pBluescript derived constructs were linearised with Asp718; pET22B derived constructs were linearised with Tth111I (see "plasmids" below). Template DNA was phenol extracted, ethanol precipitated and resuspended at 1μg/μl in DEPC treated water. For RNA synthesis, 2μg of DNA template was mixed with 4μl 5X transcription buffer (200mM Tris-HCl pH8, 40mM $MgCl_2$, 10mM spermidine, 250mM NaCl, 50mM DTT), 2μl 3.3mM ATP, GTP, TTP and CTP, 0.5μl RNAsin, 0.2μl 10mg/ml acetylated BSA and 20 units T3 or T7 RNA polymerase as appropriate, and made up to 20μl total volume with DEPC treated water. Reactions were incubated at 37°C for 30 mins, followed by addition of 2.6μl DNAse buffer and 4μl DNAseI (Promega), and further incubation at 37°C for 15 mins. RNA was extracted once with phenol/ chloroform, once

with chloroform, and precipitated by addition of 1/10 volume of 7.5M NH_4Ac and 2.5 volumes of ethanol. Pellets were washed with 80% ethanol and resuspended in DEPC treated water at 0.2mg/ml.

2.6.2 *In vitro* translation

In a typical *in vitro* translation reaction, 5µg of RNA was incubated in 120µl total volume, with 48µl nuclease treated rabbit reticulocyte lysate (Promega), 8µl 0.15M KAc, 8µl complete amino acid mixture (Promega), and 5µl RNAsin, for 50 minutes at 30°C. *In vitro* translated protein was frozen immediately in aliquots at -70°C, and thawed on ice prior to use. No loss of activity was detectable after a single round of freeze - thawing.

2.6.3 Quantification of *in vitro* translated proteins

For quantification of protein, small scale reactions were carried out in 12µl total volume, using amino acid mixture minus methionine (Promega), and 10nmol of ^{35}S-methionine. *In vitro* translated proteins were quantified by PhosphorImager analysis as follows: ^{35}S labelled proteins were run on SDS PAGE. A standard range of dilutions of ^{35}S methionine, each made up in 12µl mock translation mix, was also run on SDS PAGE. The protein and standard gels were dried down and exposed in a PhosphorImager cassette for the same amount of time. Assuming 100% labelling of proteins with ^{35}S methionine, the pmol of methionine in each protein preparation was calculated, and corrected for the number of methionine residues in the predicted protein sequence, thus giving an indication of the pmol of protein produced.

2.7 Dilution and storage of purified proteins

Purified FLP and Cre proteins were diluted to 500nM in 1M NaCl, 25mM TAPS pH8.0, 1mM EDTA, 20% glycerol.10µl aliquots were frozen in liquid nitrogen and stored at -70°C, and thawed on ice immediately prior to use.

2.8 DNA/ protein interactions

2.8.1 Gel mobility shift assay

Radioactively labelled substrates for DNA binding assays were made as follows: "Full site" substrates are those which contain two protein binding sites. They were prepared by digesting pBKSFRT, pBKSKRT or pBKSLoxP with XhoI and SpeI. "Mutated full site" substrates are those in which one binding site is inactivated by mutation, so that only one binding site of the target can be bound by recombinase. They were made by digesting pBKSFRTM and pBKSLoxM with XhoI and SpeI. "Half site" substrates are truncated after the first binding site. They were prepared by digesting pBKSFRT with XbaI, pBKSKRT/F with XhoI and XbaI, or pBKSLoxM with XhoI and SmaI.

All substrates were end- filled by Klenow polymerase, using ^{32}P-α-dCTP and unlabelled dATP, dGTP and dTTP. The efficiency of incorporation of labelled nucleotide was determined by spotting 0.2µl of the reaction before and after the addition of Klenow polymerase onto 0.1mm cellulose polyethyleneimine thin layer chromatography (TLC) paper. The reaction products were resolved by chromatography in 0.75M KPO_4, pH 3.5, and the proportion of total label incorporated into DNA was quantified by PhosphorImager analysis of the chromatograph. From this the percent of DNA labelled could be calculated. Substrates were subsequently electrophoresed on an 8% native polyacrylamide gel. The desired band was excised and eluted overnight in oligo elution buffer (300mM NaCl, 10mM Tris-HCl pH7, 1mM EDTA, 1% phenol). Substrates were then phenol extracted, ethanol precipitated and resuspended in oligo buffer (300mM NaCl, 10mM Tris-HCl pH7, 1mM EDTA). Final DNA concentration was calculated on the basis of the concentration of the radioactive nucleotide (with correction for percent DNA labelled if less than 100%), determined by scintillation counting.

Unlabelled substrates for competition analysis were prepared using the synthetic oligonucleotides carrying the FRT, KRT or LoxP sites described in "plasmids" below. To anneal corresponding pairs of oligonucleotides, 50 pmoles of each oligonucleotide were mixed with 5µl 10x ligase buffer in 50µl total volume. The mixture was heated to 75°C for 15 mins and then left to cool slowly to room temperature. The annealed oligonucleotides could be visualised by agarose gel electrophoresis.

DNA binding conditions varied according to the protein used. When using *in vitro* translated protein, the reaction contained 50mM Tris-HCl pH 7.5, 2mM $MgCl_2$, 30mM NaCl (for Kw recombinase) or 200mM NaCl (for FLP and Cre recombinases), 0.1mg/ml acetylated BSA, and 20 ng/µl sonicated salmon sperm DNA in 50µl volume. *In vitro* translated protein was incubated in this buffer for 15 minutes on ice prior to the addition of the radioactively labelled DNA to 0.2 -0.5nM. Reactions were then incubated at 23°C for a further 15 minutes, glycerol was added to 2% final concentration, and samples were loaded onto a 5% polyacrylamide gel and electrophoresed at 200V at 4°C in 1x TBE.

In the case of purified proteins (FLP and Cre recombinases only) DNA binding reactions were carried out in 25mM TAPS pH 8.0, 1mM EDTA, 180mM NaCl, 2mM $MgCl_2$, 0.1mg/ml acetylated BSA, 2.5 ng/µl sonicated salmon sperm DNA and 2% glycerol in 20µl volume. Typically, substrate concentration was 0.2 nM (giving substrate DNA concentration of 0.005 ng/µl). All components were mixed together and incubated for 15 minutes at 23°C, then electrophoresed as above. DNA/protein complexes were visualised by autoradiography and quantified by PhosphorImager analysis.

This basic reaction was modified for the following experiments:

(a) Competition analysis:

The radioactively labelled DNA substrate was mixed with a 2 to 100 fold molar excess of unlabelled competitor DNA in constant volume prior to addition of protein. The binding reaction was then carried out as above.

b) Equilibrium titrations

Reactions were started by addition of protein. Proteins were diluted prior to their addition to the reaction mix in 20% glycerol, 25mM TAPS pH 8.0, 1mM EDTA, 1M NaCl. The reaction mix buffer was adjusted so that the final concentrations of NaCl and glycerol were as given above. The DNA protein mixtures were incubated for 10 minutes at the appropriate temperature.

c) Kinetic experiments

For time course experiments, the binding reaction was started by addition of protein, and samples were loaded at time intervals as indicated in figure legends. To ensure constant resident time in the well for each sample, and to ensure that complexes migrated as compact bands, samples were allowed 5 seconds to sink into the well before the gel was turned on at 200V. The variation in experimental data was evaluated by repeating each experiment several times. Comparison of time course experiments in which samples were taken separately but were mixed from the same protein dilution and buffer prepaprations, and were loaded on the same gel, showed variation of less than 5%. Comparison of protein titration experiments showed variations of up to 5%. Comparison of experiments in which different protein dilutions, buffer mixtures or substrate preparations were used showed variations of up to 10%.

2.8.2 Recombination assay

Excision recombination substrates consisted of a 4kb Bsu36I/ NotI fragment of pSVPaZ (FRT targets), pSVPaK (KRT targets) or pSVPaX (LoxP targets), which was end-filled by Klenow polymerase using ^{32}P-α-dCTP and purified on a 0.6% 1X TAE agarose gel. Quantification of substrates was carried out essentially as described above for radiaoactively labelled bandshift substrates. For intermolecular recombination assays, an unlabelled circular plasmid carrying a single inverted repeat target site (pBKSFRT or pBKSLoxP) was used, and recombination products were detected by southern hybridisation. For the substrate titration experiment

shown in figure 4.15 (Chapter 4.3), unlabelled Bsu36I/ NotI digested pSVpaZ and paX were used. Recombination was detected by Southern hybridisation.

Assay conditions varied depending on the protein used. When using *in vitro* translated protein, reactions were typically carried out using 0.01 to 0.04nM substrate DNA in 20 µl volume containing 50mM Tris-HCl pH 7.5, 2mM $MgCl_2$, 0.1mg ml^{-1} acetylated BSA and 30mM NaCl (for KW recombinase) or 150 to 200mM NaCl (for FLP recombinase, FLP mutants and Cre recombinase).
Unless otherwise stated, recombination reactions were carried out using 1pmol of *in vitro* translated protein, and were incubated at 30°C (FLP and KW) or 37°C (Cre) for 60 mins. Reactions were terminated by addition of SDS to 0.05%, EDTA to 10mM and protease K to 0.5mg /ml followed by incubation at 37°C for 1 to 16 hours. The samples were then phenol/chloroform extracted and the aqueous phase was loaded onto a 0.6% agarose gel. After electrophoresis the gel was transferred to 3MM Whatman paper, and excess liquid was removed by rolling with a glass pipette between several layers of Whatman paper. The gel was then covered with Saran wrap and dried in a vacuum gel dryer for a maximum of 40 minutes. Recombination products were visualised by autoradiography and quantified by PhosphorImager analysis.

For time course experiments and experiments at different temperatures, reactions were carried out in a larger total volume, scaled up accordingly. Reaction mixtures were pre-incubated at various temperatures for 2 mins, and recombination reactions were started by the addition of protein. 10 or 20µl aliquots were taken at various time intervals, and reactions were terminated and analysed as described above.

In the case of purified proteins, reactions were typically carried out using 0.2nM substrate DNA in 20µl volume containing 25mM TAPS pH 8.0, 180mM NaCl, 2mM $MgCl_2$, 1mM EDTA, 20% glycerol, 10% PEG 8,000 and 2.5mg/ml BSA. Reactions were incubated, terminated and analysed as described above.

2.8.3 Strand exchange assay

The strand exchange assay used to determine the point of cleavage by KW recombinase is based on intermolecular recombination between the unlabelled circular substrate pBKSKRT, and a linear oligonucleotide substrate carrying a single inverted repeat Kw target site, which is labelled at selected positions. The synthesis of oligonucleotide substrates labelled at selected positions was carried out as follows: Two pairs of synthetic oligonucleotides were made, containing the following sequences:

1) ("Top strand" substrates)

(51 mer) 5'----CGACGAAAAATGGTAAGGA

(104mer) 3'----GCTGCTTTTTACCATTCCTTATCTGGTAAGGAATGGTAAAAAGCAGC--5'

2) ("Bottom strand" substrates)

(44 mer) 5'—CGACGAAAAATGGTAAGGA

(104 mer) 3'—GCTGCTTTTTACCATTCCTTACCAGATAAGGAATGGTAAAAAGCAGC--5'

For the labelling reaction, 5pmol of each oligo pair was mixed with 6.67 pmol of the desired ^{32}P-α-deoxynucleoside triphosphate in PCR buffer (Cetus). If the labelled base to be incorporated was an adenine, then no other dNTP was added to this reaction. If the labelled base to be incorporated was, for instance, a thymine, then unlabelled dATP was also provided. As a control, a reaction was set up with ^{32}P-α-dCTP, and no other dNTP. A full list of labelled substrates is given in figure (3.4). The labelling reaction was carried out using Amplitaq^{TM} DNA polymerase, in a single PCR cycle with 1 min denaturation at 94^{o}C, 3 mins annealing at 62^{o}C in the case of "top strand" substrates and 3 mins at 55^{o}C in the case of "bottom strand" substrates, followed by 3 mins 30 seconds elongation at 72^{o}C. Incorporation of label was checked by thin layer chromatography (20 to 50% for all "legitimate" nucleotides; less than 1% in the case of the dCTP control). Unincorporated dNTPs were removed by precipitation. The DNA was resuspended in PCR buffer, with 0.4mM each of all four unlabelled dNTPs, and the remainder of the second strand was synthesised using Amplitaq^{TM} polymerase in a single PCR cycle essentially the same as the first reaction except

that an elongation step of 5 mins was used. The labelled, double stranded oligos thus synthesised were gel purified as described in section 2.8.1 and used as substrates for recombination.

Typically, intermolecular recombination assays were carried out in a volume of 100µl, in the standard *in vitro* recombination buffer. 75fmol of a labelled oligonucleotide target site and 75fmol (= 150ng) of pBKSKRT were incubated with 10 pmol of *in vitro* translated Kw protein for 60 mins at 30°C. Reactions were stopped as described above. After phenol/ chloroform extraction the DNA was ethanol precipitated, and resuspended in ScaI digestion buffer. Half of each sample was retained for analysis of recombination products, whilst the other half was digested with ScaI. Recombination products and digestion products were run on a 0.8% agarose gel and were visualised by autoradiography.

2.9 Plasmids

2.9.1 pBluescript derived plasmids containing DNA binding targets

pBKSFRT, pBKSKRT, pBKSKRT/F, pBKSLoxP and pBKSLoxM were made by ligating annealed pairs of synthetic oligonucleotides with the following sequences (5´- 3´) into pBluescript KS+ (Stratagene) cut with HindIII and BamHI.

pBKSFRT:
GATCCCGGAAACGCTTTCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTT
CGATCGA annealed to
AGCTTCGATCGAAGTTCCTATACTTTCTAGAGAATAG
GAACTTCGAAAGCGTTTCCGG

pBKSFRTM
GATCCCGGAAACGCTTTCGAGAGGTTTATTCTCTAGAAAGTATAGGAACTT
CGATCGA annealed to
AGCTTCGATCGAAGTTCCTATACTTTCTAGAGAATAA
ACCTCTCGAAAGCGTTTCCGG

pBKSKRT:
GATCCCGGAAATCGACGACGAAAAATGGTAAGGAATAGACCATTCCTTACC
ATTTTTCGTCGATCGA annealed to
AGCTTCGACGAAAAATGGTAAGGAATG
GTCTATTCCTTACCATTTTTCGTCGTCGATTTCCGG

pBKSKRT/F:
GATCCCGGAAATCGACGACGAAAAATGGTAAGGAATCTAGAAATTCCTTAC
CATTTTTCGTCGATCGA annealed to
AGCTTCGACGAAAAATGGTAAGGAATT
TCTAGATTCCTTACCATTTTTCGTCGTCGATTTCCGG

pBKSLoxP:
GATCCCGGAAAGTACTATAACTTCGTATAATGTATGCTATACGAAGTTATC
GATCGA annealed to
AGCTTCGATCGATAACTTCGTATAGCATACATTATACG
AAGTTATAGTACTTTCCGG

pBKSLoxM:
GATCCCGGAAAGTACTATAACTTCGTATAACCCGGGCCACACGAGCTTATC
GATCGA annealed to
AGCTTCGATCGATAAGCTCGTGTGGCCCGGGTTATACG
AAGTTATAGTACTTTCCGG

2.9.2 pSV series containing recombination targets

The plasmids used as substrates for excision recombination pSVpaZ11 (FLP substrate) pSVpaK (KW substrate) and pSVpaX (Cre target) were constructed by P.-O. Angrand and are described in (Buchholz et. al 1996) and (Ringrose et.al., 1997).

2.9.3 FRED11 series containing recombination targets

The FRED11 series of plasmids used as substrates for FLP mediated excision recombination, having different distances between the FRT target sites, were constructed by Sophie Chabanis and will be

desribed in detail elsewhere. FRED11 has 74 bp between its FRT sites (measured between the centres of the spacers). A 15kb yeast DNA fragment was inserted between the FRT sites of FRED11, and a series of smaller inserts, ranging from 9.9 kb to 200 bp, were generated by restriction digestion or by exonuclease III deletion.

2.9.4 Expression vectors

a) FLP recombinase expression vectors

pBKS FLP F70L. The commercially available plasmid pOG44 (Sratagene) contains a point mutation in the FLP recombinase coding region, which changes FLP amino acid 70 from Phe to Leu. pBKS FLP F70L contains the coding sequence of FLP isolated from pOG44 as an Asp718/ PstI partial fragment, ligated into pBluescriptKSII+ (Stratagene) downstream of the T7 promoter.

pET22BFLP F70L contains a PCR generated FLP coding sequence made using pBKS FLP F70L as a template. The FLP coding sequence was ligated into pET22B+ (Novagen) such that the plasmid contains the entire FLP F70L recombinase coding sequence downstream of the T7 promoter, fused at its C- terminus to the amino acid sequence : A A A L E H H H H H H .

pET22BFLP (wild type) was made by ligating a Bsu36I/NsiI fragment of pMJ1 (Govind and Jayaram, 1987) into pET22BFLP F70L cut with the same enzymes, thus converting the FLP coding region to the wild type sequence.

pET22BFLP R258Q was made by Frank Buchholz by introducing a point mutation into the coding sequence of pET22BFLP (wild type) which converts amino acid 258 from Arg to Gln.

b) Kw recombinase expression vectors

pBSKw contains the coding sequence of the Kw recombinase isolated as a BspEI/ ScaI partial fragment from plasmid pKWS1 (Chen et al

, 1992a) ligated into pBluescribe (Stratagene) downstream of the T3 promoter.

pET22BKw contains a PCR generated Kw coding sequence, made using plasmid pBSKw as a template, ligated into pET22B+ (Novagen) such that the plasmid contains the entire Kw recombinase coding sequence downstream of the T7 promoter, fused at its C- terminus to the amino acid sequence : A A A L E H H H H H H .

c) Cre recombinase expression vector

pET22BCre contains a Cre coding sequence which was PCR amplified from plasmid pcltsCRE (Invitrogen), ligated into pET22B+ (Novagen) such that the plasmid contains the entire Cre coding sequence downstream of the T7 promoter, fused at its C- terminus to the amino acid sequence : A A A L E H H H H H H .

2.10 Mathematical methods

2.10.1 Determination of equilibrium constants for DNA binding

In the DNA mobility shift assay it is possible to distinguish between DNA molecules with an exact number (n) of monomers bound. The possible configurations for the "full site" system, composed of two linked binding sites, each of which can be bound by a recombinase monomer, are as follows:

S= free substrate (n=0)
SM= substrate with one monomer bound (n=1)
SM2= substrate with two monomers bound (n=2).

The fraction, Y, of DNA molecules in each configuration was calculated as: $Y = (S^*)/(\Sigma S^*)$, where S* = PhosphorImage counts in the band of interest, and ΣS^* = sum of PhosphorImage counts in all bands in the lane. In experiments where bound DNA molecules are expressed as nM concentrations, this value is calculated as

nM bound = Y. (nM input DNA). The binding configurations and their corresponding equilibrium association constants are shown in Table 2.1.

Table 2.1. Configurations and equilibrium constants for recombinase binding to a DNA full site substrate
Binding configurations are denoted by M if occupied.

Species	Binding configurations Site a	Site b	Equilibrium constant
S	-	-	-
SMa	M	-	K_{1a}
SMb	-	M	K_{1b}
SM2	M	M	K_2

Using the mobility shift assay with an inverted repeat substrate it is not possible to distinguish between the species SMa and SMb. The observed species SM, is equal to SMa + SMb, and has the corresponding equilibrium constant K_1, where

$$K_1 = (K_{1a} + K_{1b}) \qquad 1$$

In Table 2.1 the equilibrium constant K_2 describes the binding of a second monomer to either species SMa or SMb. Thus in a system in which the intrinsic affinities for half sites a and b are not identical, the constant K_2 will either be equal to K_{2a} or to K_{2b}, depending on which site is occupied first. For the purposes of the determination of DNA association equilibrium constants, the assumption was made that the intrinsic affinities for sites a and b are identical, thus $K_{1a} = K_{1b} = K_1/2$; and $K_{2a} = K_{2b} = K_2$.

In the full site system shown in table 2.1, the binding equations for each of the configurations are given by equations 2a to 2c, in which the terms S, SM and SM2 correspond to the fraction of total substrate in each of the three configurations:

:

$$S = \frac{1}{1+K_1.[M] + (K_1.K_2).[M]^2} \qquad 2(a)$$

$$SM = \frac{K_1.[M]}{1+K_1.[M] + (K_1.K_2).[M]^2} \qquad 2(b)$$

$$SM2 = \frac{(K_1.K_2).[M]^2}{1+K_1.[M] + (K_1.K_2).[M]^2} \qquad 2(c)$$

(Senear and Brenowitz, 1991). S, SM and SM2 are the fractions of total DNA in each of the configurations. [M] is the concentration of free protein monomer. Titrations of FLP and Cre were performed against a fixed amount of full site DNA substrate, and the fraction of DNA in each species was plotted against the concentration of free monomer. The equations 2 (a) to (c) were then simultaneously fitted to the data to obtain values for K_1 and K_2.

In order to determine the individual constants K1a and K1b, the mutated target sites FRTM and LoxM were used, in which only one site can be bound by recombinase monomer. The possible configurations in this system are S, and SMa. The equation for binding to a single site is given by the Langmuir isotherm (Senear and Brenowitz, 1991):

$$SMa = \frac{K_{1a}.[M]}{1+K_{1a}.[M]} \qquad 3$$

Where SMa is the fraction of total DNA bound. Titrations of FLP and Cre were performed against a fixed amount of FRTM or LoxM DNA substrate, and the fraction of DNA in SMa was plotted against the concentration of free monomer. Equation 3 was fitted to the data using the Kaleidagraph program (Albeck software) for the Apple Mackintosh computer, to obtain the value of K_{1a}. The value of K_{1b} is then given by equation 1; $K_1=(K_{1a}+K_{1b})$.

In all cases except Cre binding to LoxP full site, the assumption was made that $[M]_{total} = [M]_{free}$, because the concentration of bound DNA at each protein input was low (less than 2%) relative to the total protein concentration. In the case of the Cre/LoxP experiments, $[M]_{free}$ was calculated as

$$[M]_{free} = [M]_{total} - (2[SM2]+[SM]) \qquad 4$$

(In equation 4, SM and SM2 are expressed in terms of their molar concentration). Accurate determination of SM2 and SM using equation 4 is dependent on accurate determination of the substrate concentration, since [SM1] and [SM] are determined as (fraction bound).(total input DNA). It was estimated that the variation in quantification of independently prepared substrates is up to 10%. Therefore in the case of Cre/LoxP this introduces a potential source of additional error to the determination of K_1 and K_2 using the full site system.

2.10.2 Estimation of rate constants for DNA binding

Approximate values of the rate constants for DNA binding and dissociation were calculated individually as descibed below. These values were further refined by fitting simulated DNA binding data to experimental time course curves as described in section 2.10.3.
The binding of recombinase monomers to a full site substrate is described by the kinetic equations:

$$S + M \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} SM \qquad 5$$

$$SM + M \underset{k_{-2}}{\overset{k_2}{<\!\!-\!\!-\!\!>}} SM2 \qquad 6$$

k_1 and k_{-1} are the association and dissociation rate constants respectively, for the species SM. k_2 and k_{-2} are the association and dissociation rate constants for the species SM2.

The evolution of species SM and SM2 over time is described by the differential equations:

$$d[SM]/dt = k_1[S][M] - k_{-1}[SM] - (k_2[SM][M] - k_{-2}[SM2]) \qquad 7$$

$$d[SM2]/dt = k_2[SM][M] - k_{-2}[SM2] \qquad 8$$

d[SM] and d[SM2] are the change in molar concentration of species SM and SM2 respectively. dt is the time interval in seconds. The terms in square brackets refer to molar concentrations at time t. [S] is the concentration of free substrate; [M] is the concentration of free monomer. For all time course experiments, [M] was calculated according to equation 4 above.

Association rate constant, k_{1a}

For a single site system containing only one binding site (a), the evolution of species SMa is described by the equation:

$$d[SMa]/dt = k_{1a}[S][M] - k_{-1a}[SM] \qquad 9$$

Where k_{1a} and k_{-1a} are the association and dissociation rate constants respectively for binding to site a.

The rate constant k_{1a} was determined using the single site system (mutated full sites and half sites) for both FLP and Cre as follows:

At the start of the binding reaction, [SM] is small, and the back rate, described by the term k_{-1a} [SM] in equation 9, is minimal.

The approximate rate constant k_{1a} is thus given by the equation:

$$k_{1a} = \frac{d[SMa]}{[S][M].dt} \qquad 10$$

Rate constants were determined based on measurements taken at the earliest time point, 15 seconds.

Apparent assciation rate constant k_1

In the full site system, an apparent value for the rate constant k_1 was calculated according to the following reasoning:

The quantity of interest is the quantity of SM which would be present if it were not consumed by formation of SM2. This theoretical value, "SM*" can be used to give an indication of the apparent rate constant k_1 ("k_{1*}") in the full site system.

Substituting equation 8 in equation 7 gives:

$$d[SM]/dt = k_1[S][M] - k_{-1}[SM] - (d[SM2]/dt) \qquad 11$$

The term (d[SM2]/dt) accounts for those SM complexes which are consumed into complex SM2, resulting in a reduction in the quantity of SM. The term $k_1[S][M] - k_{-1}[SM]$ describes the new formation of SM over the time interval, t, which is the quantity of interest, SM*. Substituting the term d[SM*]/dt, into equation 11 gives:

$$d[SM]/dt = d[SM^*]/dt - d[SM2]/dt$$

$$\text{and } d[SM^*]/dt = d[SM]/dt + d[SM2]/dt. \qquad 12$$

Thus by adding together the values of [SM] and [SM2] observed at the15 second time point, a theoretical value for [SM*] is obtained.

The apparent rate constant, k_1*, for the formation of this species was calculated using equation 10.

Apparent association rate constant k_2.

In a full site system, the evolution of the species SM2 is described by equation 8. At the start of the binding reaction, where the concentration of SM2 is low, the back rate described by the term $k_{-2}[SM2]$ is minimal. Thus the approximate rate constant, k_2, is given by the equation:

$$k_2 = \frac{d[SM2]}{[SM][M].dt} \qquad 13$$

k_2 was determined using data from the 15 second time point.

Dissociation rate constants k_{-1}, k_{-1a} and k_{-2}

Dissociation constants k_{-1}, k_{-1a} and k_{-2} (in general: k_{-n}) were calculated using the measured equilibrium constants (K_n) (section 2.10.1) and corresponding estimated association rate constants (k_n) (this section, above) according to the relationship:

$$K_n = \frac{k_n}{k_{-n}} \qquad 14$$

2.10.3 Mathematical modelling I: DNA binding

Equations 7 and 8, describe the rate of change in concentration of the species SM and SM2 with respect to time, in terms of the species from which they are formed, the species into which they can be converted, and the rate constants which govern these processes. For a full mathematical description of the system,

differential equations describing the behaviour of the free monomer (M) , and of the free substrate (S) are also required:

$$d[M]/dt = -k_1[S][M] + k_{-1}[SM] - k_2[SM][M] + k_{-2}[SM2] \quad 15$$

$$d[S]/dt = -k_1[S][M] + k_{-1}[SM] \quad 16$$

Equations 7, 8, 15 and 16 were implemented in a Fortran 77 code, and solved by finite difference numerical integration.

2.10.4 Mathematical modelling II: recombination.

The scheme upon which the mathematical model of recombination is based is shown in figure 4.1. The species described by the model are listed in Table 2.2, and are diagrammed in Table 4.4.
The interconversion of the species in Table 2.2 is described by the kinetic equations:

$$S + M \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} SM \quad 5$$

$$SM + M \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} SM2B \quad 6$$

$$SM + M \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} SM2A \quad 17$$

Table 2.2 Species in mathematical model for recombination.

Name	Description	Rate constants
M	free monomer	
S	free substrate	
SM	one monomer bound	
SM2A	one monomer at each site	k_1, k_{-1}
SM2B	two monomers at one site	k_2, k_{-2}
SM3	three monomers bound	
SM4	four monomers bound	
I	synaptic complex	k_3, k_{-3}
IEP	excised synaptic complex	k_4, k_{-4}
EPM2	linear excision product with two monomers bound	k_5, k_{-5}
LPM2	circular excision product with two monomers bound	k_5, k_{-5}
EPM	linear excision product with one monomer bound	k_{-2}, k_2
LPM	circular excision product with one monomer bound	k_{-2}, k_2
EP	free linear product	k_{-1}, k_1
LP	free circle	k_{-1}, k_1

$$SM2B + M \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} SM3 \qquad 18$$

$$SM2A + M \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} SM3 \qquad 19$$

$$SM3 + M \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} SM4 \qquad 20$$

$$SM4 \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} I \qquad 21$$

$$I \underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} IEP \qquad 22$$

$$IEP \underset{k_{-5}}{\overset{k_5}{\rightleftharpoons}} EPM2 + LPM2 \qquad 23$$

$$EPM2 \underset{}{\overset{k_{-2}}{\rightleftharpoons}} EPM + M \qquad 24$$

$$LPM2 \underset{k_2}{\overset{k_{-2}}{\rightleftharpoons}} LPM + M \qquad 25$$

$$EPM \underset{k_1}{\overset{k_{-1}}{\rightleftharpoons}} EP + M \qquad 26$$

$$LPM \underset{k_1}{\overset{k_{-1}}{\rightleftharpoons}} EP + M \qquad 27$$

From equations 5, 6, and 17 to 27, and using classical kinetic principles, the following differential equations, describing the evolution of each species over time, were derived:

$$d[S]/dt = -k_1[S][M] + k_{-1}[SM] \qquad 28$$

$$\begin{aligned} d[M]/dt = {} & -k_1[S][M] + k_{-1}[SM] - k_2[SM][M] + k_{-2}[SM2B] \\ & - k_1[SM][M] + k_{-1}[SM2A] - k_2[SM2A][M] + k_{-2}[SM3] \\ & - k_1[SM2B][M] + k_{-1}[SM3] - k_2[SM3][M] + k_{-2}[SM4] \\ & - k_2[EPM][M] + k_{-2}[EPM2] - k_2[LPM][M] + k_{-2}[LPM2] \\ & - k_1[EP][M] + k_{-1}[EPM] - k_1[LP][M] + k_{-1}[LPM] \end{aligned} \qquad 29$$

$$\begin{aligned} d[SM]/dt = {} & k_1[S][M] - k_{-1}[SM] + k_{-1}[SM2A] - k_1[SM][M] \\ & + k_{-2}[SM2B] - k_2[SM][M] \end{aligned} \qquad 30$$

$$d[SM2A]/dt = k_1[SM][M] - k_{-1}[SM2A] + k_{-2}[SM3] - k_2[SM2A][M] \qquad 31$$

$$d[SM2B]/dt = k_2[SM][M] - k_{-2}[SM2B] + k_{-1}[SM3] - k_1[SM2B][M] \quad 32$$

$$d[SM3]/dt = k_1[SM2B][M] - k_{-1}[SM3] + k_2[SM2A][M] - k_{-2}[SM3] + k_{-2}[SM4] - k_2[SM3][M] \quad 33$$

$$d[SM4]/dt = k_2[SM3][M] - k_{-2}[SM4] + k_{-3}[I] - k_3[SM4] \quad 34$$

$$d[I]/dt = k_3[SM4] - k_{-3}[I] + k_{-4}[IEP] - k_4[I] \quad 35$$

$$d[IEP]/dt = k_4[I] - k_{-4}[IEP] + k_{-5}[EPM2][LPM2] - k_5[IEP] \quad 36$$

$$d[EPM2]/dt = k_5[IEP] - k_{-5}[EPM2][LPM2] + k_2[EPM][M] - k_{-2}[EPM2] \quad 37$$

$$d[LPM2]/dt = k_5[IEP] - k_{-5}[EPM2][LPM2] + k_2[LPM][M] - k_{-2}[LPM2] \quad 38$$

$$d[EPM]/dt = k_1[EP][M] - k_{-1}[EPM] + k_{-2}[EPM2] - k_2[EPM][M] \quad 39$$

$$d[LPM]/dt = k_1[LP][M] - k_{-1}[LPM] + k_{-2}[LPM2] - k_2[LPM][M] \quad 40$$

$$d[EP]/dt = k_{-1}[EPM] - k_1[EP][M] \quad 41$$

$$d[LP]/dt = k_{-1}[LPM] - k_1[LP][M] \quad 42$$

equations 28 to 42 were implemented and solved as described above. The input values of all rate constants, and of initial protein and substrate concentrations can be varied. The quantity taken to represent the total excision product, EPtot, is given by IEP + EPM2 + EPM + EP. This quantity is considered to describe the experimentally observed excision product because, firstly, only the linear excision product, and not the circle, is labelled and quantified. Secondly, because recombination reactions are terminated with protease K, the DNA seen as excision product includes the DNA within the

excised synaptic complex (IEP in the model) and all subsequent bound and unbound forms (EPM2, EPM, and EP).

CHAPTER 3
Characterisation of Kw recombinase, a new integrase from *Kluyveromyces waltii.*

3.1 Introduction

The integrase family of site specific recombinases share limited amino acid sequence homology, but use a common reaction mechanism to recombine distinct target sites. (Argos et al., 1986, Abremski and Hoess, 1992, Figure 1.1, Chapter 1.2). A further subfamily within the integrase family is the yeast family, whose members share somewhat higher amino acid sequence homology (Utatsu et al., 1987, Figure 1.1). The best characterised member of this family is the FLP site specific recombinase, encoded on the 2μ circle from *S. cerevisiae* (reviewed by Sadowski, 1995, and Chapter 1). By analogy to the FLP system, other yeast family recombinases and their recombination target sites have been identified from the 2μ like plasmids of other yeast species (Utatsu et al., 1987, Araki et al., 1992, Chen et al., 1986, Chen et al., 1992a).

The simplicity of these recombination systems means that they are amenable to studies of the recombination reaction itself, certain features of which are not yet fully understood (for reviews see Sadowski, 1995, and Stark and Boocock, 1995). The presence in yeasts of both the recombinase coding sequence and the DNA target sequences on the same plasmid, facilitates the identification of new members, whose characterisation would add to the understanding of the details of site specific recombination, and of the relationships among the yeast and integrase families.

The characterisation of new site specific recombinases is desirable not only in the context of comparative studies, but also because of their potential application to the field of genome engineering. The yeast site specific recombinases FLP and R (Araki et al., 1992) and the site specific recombinase Cre, from bacteriophage P1, have been used to direct DNA rearrangements in living systems (reviewed in Kilby et al., 1993, and chapter 1.5). The utility of these recombinases rests on certain shared features, namely: 1) The reaction is carried out by a single protein. 2) The minimal target site required for efficient recombination is quite simple, consisting of two 12 or 13 base pair repeats in inverted orientation, separated by a non palindromic spacer of 7 or 8 base pairs. 3) The outcome of recombination between two DNA target sites is dictated by the

orientation of the spacers with respect to one another. 4) These recombinases catalyse both intramolecular recombination (excision or inversion) and intermolecular recombination (integration or translocation). 5) Recombination is conservative, with no net loss or gain of nucleotides. 6) All three recombinases function in a range of different host organisms.

The identification of novel recombinases with distinct target site specificity, which share the features listed above, would widen the repertoire of tools available for the manipulation of living genomes. This chapter describes the characterisation of a novel member of the yeast family of site specific recombinases, the "Kw" recombinase, identified from the pKW1 plasmid of the yeast *Kluyveromyces waltii.* (Chen et al., 1992a). The results presented here show that the Kw and FLP proteins have distinct target site specificity. The Kw protein is alone sufficient to promote both inter- and intramolecular recombination *in vitro* of a DNA target site consisting of two 18 base pair elements separated by a 7 base pair spacer. Recombination is conservative. In addition, Kw recombinase functions in mammalian cells in culture. The results are discussed both in the context of comparison of Kw recombinase with other yeast family members, and in connection with its potential use as a tool for genomic manipulation.

3.2 Results

3.2.1 Kw binds specifically to its target site.

The 2μ–like plasmid, pKW1 from *Kluyveromyces waltii* contains four open reading frames, A, B, C and D. Chen et al (1992a) suggested that the 'A' gene encodes a recombinase, on the basis of homology of its deduced amino acid sequence with that of the FLP gene of 2μ and its equivalent in other yeast plasmids.

Like other 2μ plasmids, pKW1 also has two inverted repeat sequences of about 300 bp. Within this region, the putative recombination target sites have been identified by Chen et. al. (1992a) on the basis of their sequence structure. Each putative target site contains three 18 base pair perfect repeat elements, two of which are inverted, separated by a non palindromic stretch of 7 base pairs. The third repeat element is in direct orientation to the second, and is separated from it by 26 base pairs. This is reminiscent of the situation in the *S. cerevisiae* 2μ plasmid, in which the FLP recombination target region also contains an inverted repeat with a single adjacent direct repeat element (Broach et al., 1982). With FLP, however, the direct repeat is separated from the inverted repeat by 2 base pairs. The inverted repeat of two 13 base pair elements is sufficient for efficient FLP recombination *in vitro* (Senecoff et al., 1985) and *in vivo* (Jayaram, 1985).

To determine whether the putative Kw recombinase would bind its putative DNA target site, *in vitro* translated Kw protein was tested in a gel mobility shift assay.

In vitro translation of Kw protein from both pBSKw and pET22BKw gave a protein of approximately 45kDa determined by SDS PAGE, which corresponds to the molecular weight predicted from the amino acid sequence (data not shown).

Figure 3.1 Binding of *in vitro* translated Kw recombinase to its DNA target site. a) Sequence of DNA target site oligos used for gel mobility shift assay. Preparation of [^{32}P] -labelled oligos is described in "Materials and Methods". *Open boxes* indicate repeat elements of the Kw and FRT target sites. **b) Gel mobility shift assay.** In vitro translated Kw protein binding to target sites carrying one or two repeat elements. Lanes 1-4: 2 fmol of full site oligo was incubated with 2, 4, 8, or 16 µl Kw protein translated *in vitro* from pBSKW. (0.83 pmol/µl) Lanes 5-8: 2 fmol half site oligo was incubated with 2, 4, 8, or 16µl Kw protein. Lane 9: 8 µl Kw protein incubated with 2 fmol FRT site oligo. Lane 10: 8 µl mock translation with 2 fmol KW half site. **c) Competition with unlabelled oligo target sites.** 4µl Kw protein was bound to 2 fmol full site labelled oligo in the presence of an excess of unlabelled Kw full site oligo. Lane 1: mock translation; lane 2: Kw, no competitor; lanes 3-6: 5, 10, 20, 50 fold molar excess of competitor DNA. Lane 7: 50 fold molar excess of unlabelled FRT site oligo.

Figure 3.1

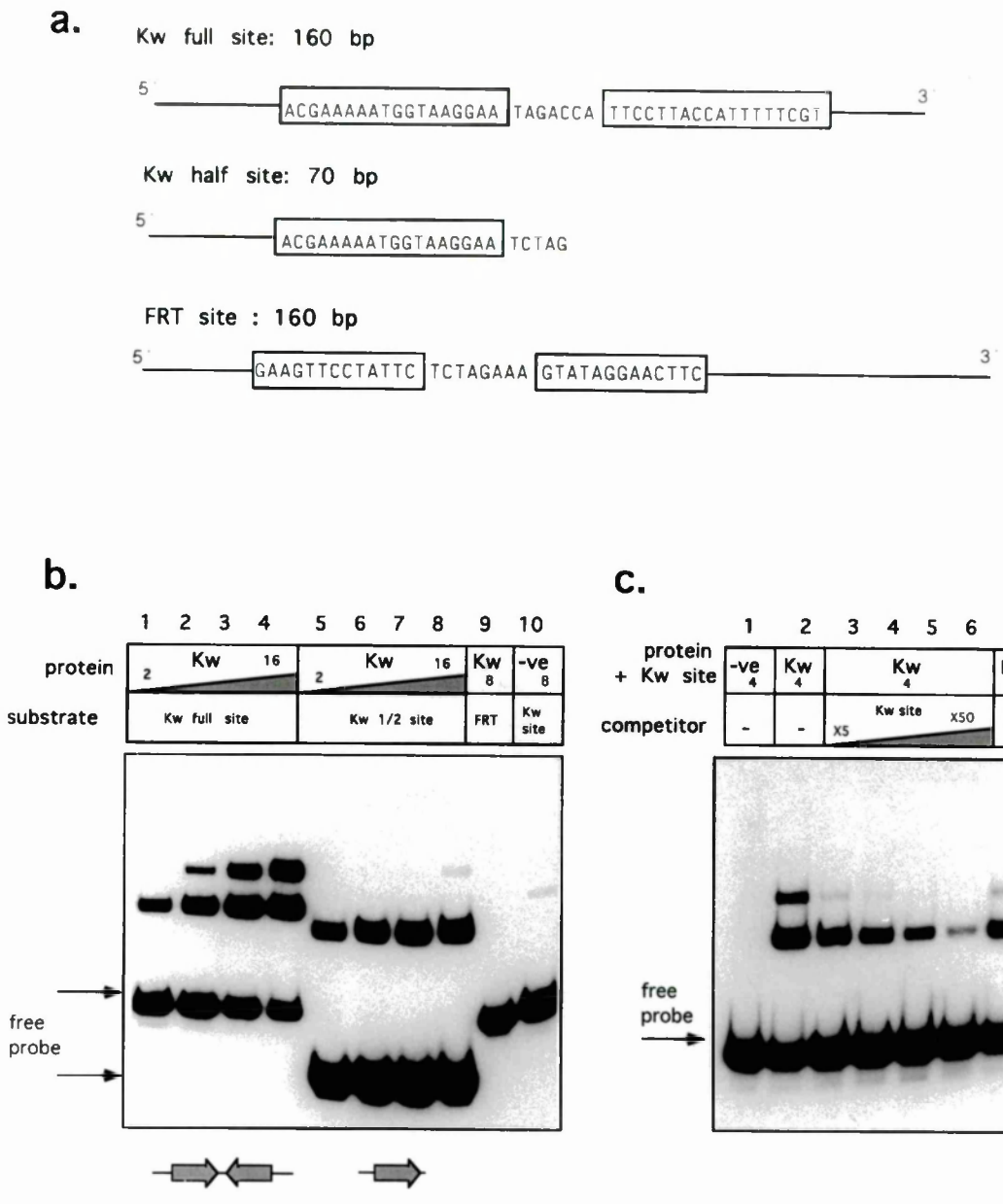

a.
Kw full site: 160 bp
5'
ACGAAAAATGGTAAGGAA
TAGACCA
TTCCTTACCATTTTTCGT
3'
Kw half site: 70 bp
5'
ACGAAAAATGGTAAGGAA
TCTAG
FRT site : 160 bp
5'
GAAGTTCCTATTC
TCTAGAAA
GTATAGGAACTTC
3'
b.
1 2 3 4 5 6 7 8 9 10
protein
substrate
Kw
2
16
Kw full site
Kw 1/2 site
FRT
Kw site
-ve
8
free probe
c.
1 2 3 4 5 6 7
protein + Kw site
competitor
-ve
Kw
4
X5
X50
FRT
free probe

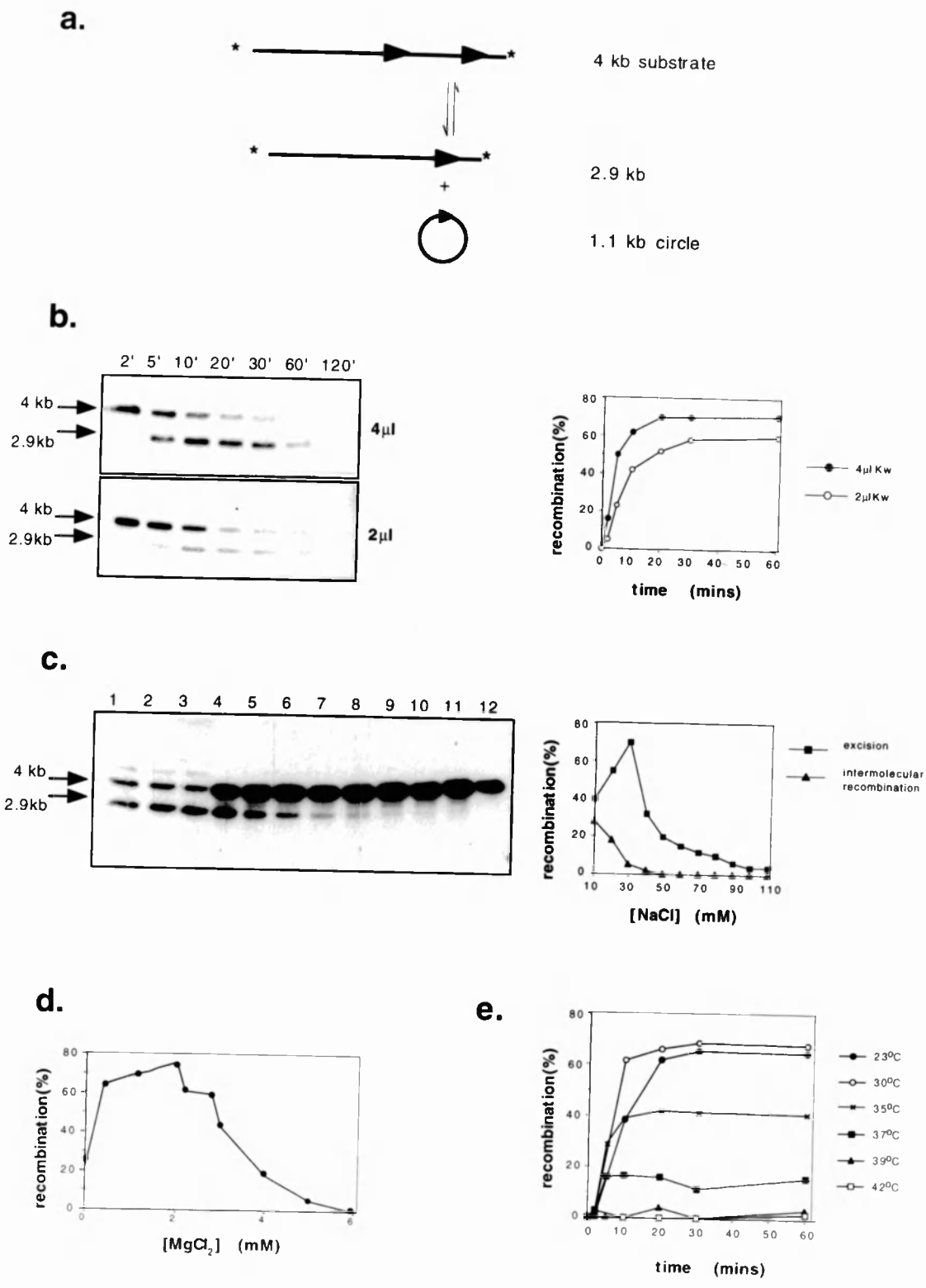
Figure 3.1a shows the sequence of DNA target sites used for the bandshift assay. Kw target sites containing two repeat elements in inverted orientation (Kw full site), as well as target sites containing a single element (Kw half site) were tested. Fig 3.1b shows that *in vitro* translated Kw protein forms two distinct complexes on a full site target. The formation of the upper complex occurs at higher protein concentrations. These two complexes probably correspond to occupation by Kw protein of one or both halves of the target site, as has been shown to be the case for FLP (Andrews et al., 1987). Kw protein is also able to bind to a half site target. (fig 3.1b lanes 5-8) Interestingly, at the highest protein input (lane 8) a faint band is visible which comigrates with the upper complex in the full site assay. This band may correspond to a complex consisting of two Kw monomers, each bound to a single DNA site. Similar complexes of FLP monomers bound to half site targets have been observed (Qian et al., 1990, Serre and Jayaram, 1992).

Fig 3.1b, lane 9 also demonstrates that Kw protein does not bind the FRT target site in this assay. The specificity of binding was confirmed in a competition experiment, shown in fig 3.1c. This experiment demonstrates that an excess of unlabelled Kw full site specifically competes for Kw protein binding. With increasing concentration of competitor DNA, from a 5 fold to a 50 fold excess, first the upper complex, and then the lower complex, is lost. Kw binding to its site is not competed by a 50 fold excess of the FRT site. In a similar experiment with *in vitro* translated FLP protein, FLP does not bind the Kw target site, and binding is competed out only by the FRT (data not shown).

These results demonstrate that Kw protein specifically binds its putative recombination target site, and suggest that the stoichiometry of its binding is identical to that shown for FLP.

Figure 3.2 Kw mediated *In vitro* recombination. a) Recombination substrates and products. Preparation of end -filled labelled recombination substrates is described in "Materials and Methods". See text for description of assay. **b) Recombination time course.** 2µl (plus 2µl mock translation) or 4µl Kw protein translated *in vitro* from pET22BKW (1 pmol/µl) was incubated with 3.2fmol substrate in 80µl reaction volume. 10µl aliquots were taken at 2, 5, 10, 20, 30 , 60 and 120 minute time intervals. **c). Effect of NaCl on recombination.** 1µl Kw protein was incubated with 0.8 fmol substrate in 20 µl for 1 hour at different NaCl concentrations, with 2mM $MgCl_2$. Lanes 1-11: 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, and 110mM NaCl . Lane 12: mock translation in 30 mM NaCl. **d) Effect of $MgCl_2$.** Experimental conditions were as for (c) above, but $MgCl_2$ concentration was varied between 0 and 6 mM. NaCl concentration was 30mM. **e) Effect of temperature on Kw recombinase activity.** Recombination reactions were carried out at 23, 30, 35, 37, 39 and 42°C. Reaction mixtures were pre warmed for 2 mins prior to the addition of protein. 6µl Kw protein translated *in vitro* from pET22BKW (0.83 pmol/µl) was added to 2.4 fmol substrate in 60µl volume, and 10 µl aliquots were taken at 2, 5, 10, 20, 30 and 60 minute time intervals.

Figure 3.2


a.
4 kb substrate
2.9 kb
+
1.1 kb circle
b.
2' 5' 10' 20' 30' 60' 120'
4 kb
2.9kb
4µl
4 kb
2.9kb
2µl
recombination(%)
time (mins)
4µl Kw
2µl Kw
c.
1 2 3 4 5 6 7 8 9 10 11 12
4 kb
2.9kb
recombination(%)
[NaCl] (mM)
excision
intermolecular recombination
d.
recombination(%)
[MgCl2] (mM)
e.
recombination(%)
time (mins)
23°C
30°C
35°C
37°C
39°C
42°C

3.2.2 Kw protein recombines its target site *In vitro*

The Kw protein was assayed for recombinase activity using a recombination substrate made from pSVpaK, which carries two Kw "full sites" arranged with their spacers in direct orientation. Kw mediated recombination of this substrate should result in excision of the 1.1 kb between the two targets. A 4kb linear end labelled fragment of pSVpaK was used for *in vitro* recombination assays. The recombination reaction is shown in fig 3.2a. Recombination gives rise to a labelled linear product of 2.9 kb and an unlabelled circle. Intermolecular recombination is also possible, either by uneven exchange between two linear substrates, or by insertion of the excised 1.1 kb circle into any linear molecule. Intermolecular recombination can be seen in this assay as molecules of 5.1 kb and larger.

Fig 3.2b shows a time course of recombination at two different protein concentrations. Clearly Kw induces recombination of its target. At a substrate concentration of 0.04nM (in 80µl volume), the amount and speed of excision is dependent on protein input: 3.32 pmol (=4 µl) of *in vitro* translated Kw gives 70% maximum recombination after 20 mins, whereas 1.66 pmol Kw (=2µl, mixed with an equal volume of mock translation so that both experiments contained equivalent amounts of lysate), gives 50% maximum recombination after 30 mins. At higher protein inputs, the excision reaches 80% (data not shown). The reaction is theoretically reversible, but it is not possible in this assay to distinguish between the unrecombined 4kb substrate, and the product of intermolecular recombination between the excised 1.1 kb circle and a 2.9 kb excision product. However, it is clear that the excision reaction is favoured over the integration reaction under these assay conditions, because bands of 5.1 kb and larger, corresponding to intermolecular recombination products, although visible, are much fainter than the 2.9 kb excision product band. These bands are more prevalent at higher substrate concentrations (data not shown).

Subsequent experiments were carried out to determine the optimum buffer conditions for the *in vitro* recombination reaction. The reaction occurs efficiently within a pH range from 6 to 8, with

optimum at pH 7.5. The recombination reaction is inhibited by TAPS buffer (data not shown).

Fig 3.2c shows the effect of NaCl concentration on Kw activity. The optimum NaCl concentration for excision is 30mM, and the reaction is strongly inhibited at concentrations above 50mM. Interestingly, the amount of intermolecular recombination observable is highest at 10mM NaCl, and is rapidly reduced as NaCl concentration is increased, being undetectable above 40mM. Fig 3.2d shows the effect of $MgCl_2$ concentration on excisive recombination. The activity is sensitive to $MgCl_2$, having an optimum at 2mM, and being completely inhibited at 6mM.

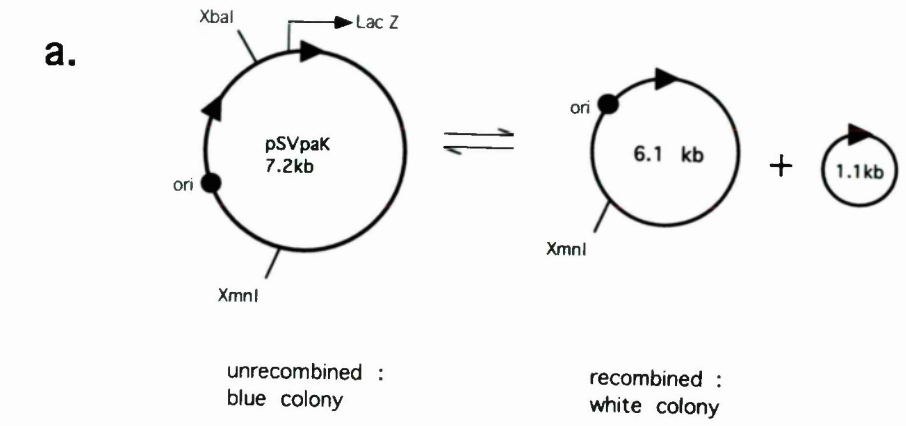
The effect of temperature on the *in vitro* recombination reaction was assessed (fig. 3.2e). Recombination was tested at 23, 30, 35, 37, 39 and 42°C. Of those tested, the optimum temperature for Kw *in vitro* recombination is 30°C, giving 67% excision after 1 hour. The reaction is increasingly inhibited at higher temperatures, with 40% and 17.6% excision after one hour at 35°C and 37°C respectively. No activity was detected at 39°C and at 42°C. However, this inhibition does not appear to be due to a slower reaction rate, since reactions at 30, 35 and 37°C show the same initial rate of excision. This result suggests that the inhibition at higher temperatures may be due to protein instability.

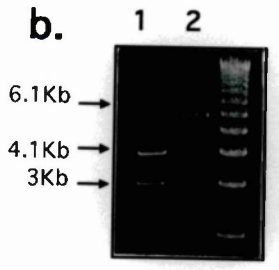
3.2.3 Kw mediated recombination is conservative

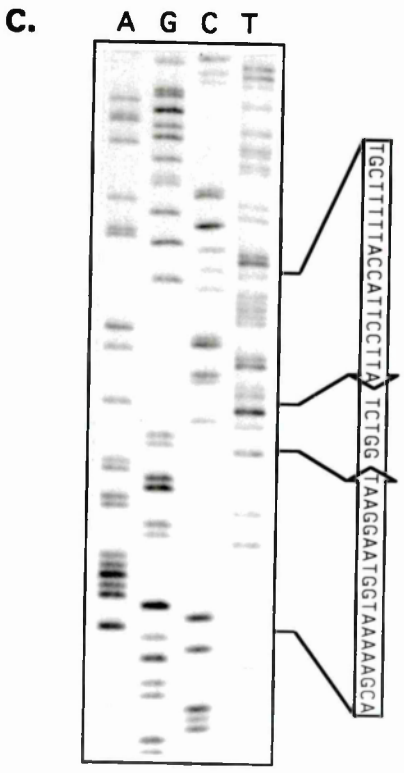
One feature of integrase class site specific recombinases is that they carry out recombination in a conservative manner: there is no net loss or gain of DNA, and the DNA target site sequence of the substrate is conserved in the product. The assay shown in fig 3.3a was designed to determine whether this is also the case for the Kw recombinase. The assay uses the pSVpaK plasmid, which carries a LacZ reporter

Figure 3.3 Conservation of Kw target site sequence upon recombination. 5µl of Kw protein translated *in vitro* from pET22BKW (1 pmol/µl) was incubated with 300ng supercoiled pSVpaK *in vitro* for 1 hour. Recombination products were transformed into *E. coli* HB101 and plated on X-gal plates for blue/ white screen. Blue colonies (unrecombined) and white colonies were picked for isolation of plasmid DNA, which was restriction mapped and sequenced. **a) Recombination substrates and products. b) XmnI/ XbaI restriction digest.** DNA from a blue colony (lane 1) and a white colony (lane 2). **c) Sequence** of DNA from a white colony.

Figure 3.3


a.
XbaI
Lac Z
pSVpaK
7.2kb
ori
XmnI
ori
6.1 kb
1.1kb
XmnI
unrecombined :
blue colony
recombined :
white colony


b.
1 2
6.1Kb
4.1Kb
3Kb


c.
A G C T
TGCTTTTTACCATTCCTTA TCTGG TAAGGAATGGTAAAAAGCA

gene whose promoter is on the DNA lying between the two Kw target sites. Upon Kw mediated recombination of pSVpaK, the promoter is excised. After *in vitro* recombination, pSVpaK was transformed into an *E. coli* LacZ- host and plated on X-gal plates. 10% of the colonies were white (recombined), and 90% were blue (unrecombined). DNA from white and from blue colonies was restriction mapped (Figure 3.3b). All white colonies picked showed the expected restriction pattern for recombined pSVpaK. Sequencing across a recombined Kw target site (Figure 3.3c) showed that the target site is conserved in the recombination product.

3.2.4 The Kw target site has a spacer of 7 nucleotides

Characterised site specific recombinases of the Int family show common features in the mechanism by which they cleave their DNA substrate (for review, see Stark et al., 1993). The catalytic tyrosine, which is absolutely conserved throughout the integrase family, makes a single strand break at a specific nucleotide at the 5' end of the spacer. The tyrosine becomes covalently attached to the 3´ phosphate at the cleavage point. It is reasonable to assume that Kw recombinase proceeds in the same way, since it belongs to the integrase family, sharing the four conserved catalytic site residues (Figure 1.1) and recombines a target site of similar sequence structure in a conservative manner.

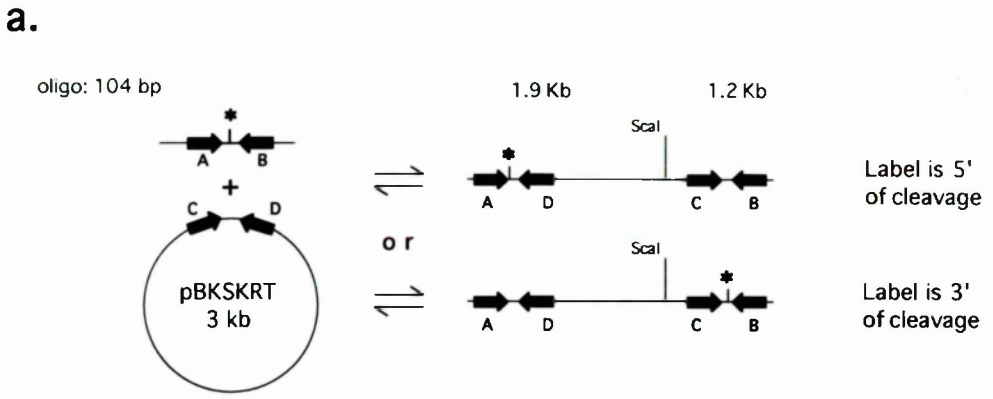
To determine which nucleotide bond is cleaved by Kw recombinase an assay was designed in which the point of strand exchange is determined by analysis of recombined products. The assay, outlined in Figure 3.4a, is based on intermolecular recombination between an unlabelled circular plasmid (pBKSKRT), and an oligonucleotide in which a phosphate at a specific position (or positions) in or adjacent to the spacer is labelled. Kw mediated recombination between the oligo and the plasmid generates a linear molecule of 3.1 Kb. As shown in figure 3.4a, if the labelled phosphate is on the 5' side of the strand cleavage point, or becomes covalently attached to the catalytic

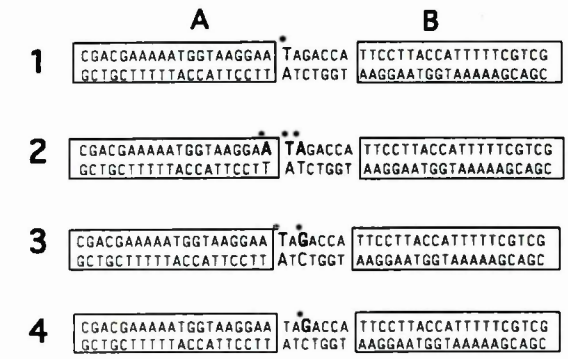
Figure 3.4 Assay to determine the point of strand cleavage by Kw recombinase.

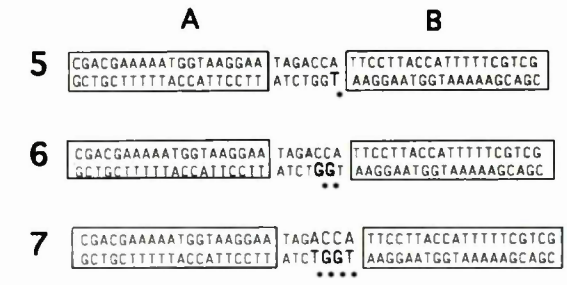
a) Recombination reaction and restriction analysis of recombination products. Arrows indicate inverted repeat elements of the Kw target site. Kw mediated intermolecular recombination between a labelled oligonucleotide substrate and an unlabelled 3kb plasmid carrying a single Kw target site ("pBKSKRT"; see "Materials and Methods") gives a single linear molecule of 3.1Kb. Restriction digest of linear recombination products with ScaI gives fragments of 1.9 Kb and 1.2 Kb (see text for full description of assay).

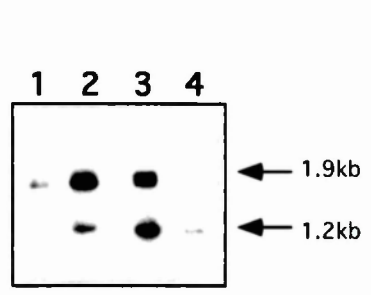
b),c) Oligonucleotide substrates. Preparation of oligonucleotide recombination substrates in which selected phosphates are labelled, is described in "Materials and Methods". Open boxes indicate the repeat elements of the Kw target site; Labelled nucleotides used in synthesis are indicated with bold type; radioactive phosphates are indicated with asterisks. **d), e) ScaI digest of recombination products.** The assay was carried out using 25 µl of Kw protein translated *in vitro* from pET22BKW (1 pmol/µl) and incubated with 75fmol of pBKSKRT and 75 fmol of each of substrates 1 to 7, in 100 µl total volume. Recombination products were digested with ScaI. Lanes 1 to 7 show ScaI digests corresponding to substrates 1 to 7.

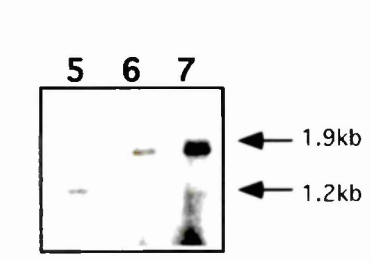
Figure 3.4


a.
oligo: 104 bp
1.9 Kb
1.2 Kb
ScaI
A
B
C
D
pBKSKRT
3 kb
or
Label is 5'
of cleavage
Label is 3'
of cleavage


b. Top strand substrates
A
B
1
2
3
4


c. Bottom strand substrates
A
B
5
6
7


d. ScaI digest
1 2 3 4
1.9kb
1.2kb


e. ScaI digest
5 6 7
1.9kb
1.2kb

tyrosine, then it will partition to the "left" side of the linear 3.1 kb molecule. If the labelled phosphate is on the 3' side of the cleavage point, then it will partition to the "right" side. Restriction digestion of linear recombination products with ScaI reveals the position of the labelled phosphate, and thus the point at which the oligonucleotide substrate was cleaved. As shown in figure 3.4a, bands of 1.9Kb and 1.2 Kb are generated.

The oligonucleotide substrates shown in figures 3.4b and c were made by labelling various positions in the Kw target site spacer on each strand by primer extension using [α-^{32}P] labelled dNTPs as described in "Materials and Methods" (section 2.8.3). Note that upon strand cleavage, when the catalytic tyrosine becomes covalently attached to a labelled 3´ phosphate, this phosphate and the base which originally carried it, will be separated.

Kw protein was incubated in separate reactions with pBKSKRT and an equimolar quantity of each of the seven labelled oligonucleotide substrates. In each case, intermolecular recombination gave a labelled linear product of 3.1 Kb (not shown). The recombination products were then digested with ScaI, shown in figures 3.4d and e. Lanes 1 to 4 of figure 3.4d correspond to substrates 1 to 4 (figure 3.4b). In lane 1, ScaI digestion yields a single labelled band of 1.9 kb. As seen in figure 3.4b, substrate no 1 is labelled at a single position, that is, the phosphate on the 5' side of the first 'T' in the spacer. This result demonstrates that this phosphate is situated on the 5' side of the point of strand exchange, as it remains attached to the "left" arm of the target site (figure 3.4a). Substrate no. 2 (figure 3.4d lane 2) yields both of the two bands, of 1.9 and 1.2 kb. In this substrate, three consecutive nucleotides: 'ATA' were labelled. Therefore the three labelled phosphates must distribute either 3:0, 2:1, 1:2 or 0:3 across the cleavage site. The relative intensities of the 1.9Kb and 1.2Kb bands indicate approximately 2:1 distribution. Since substrate no. 1 demonstrates that the phosphate from the 'T' is on the 5' side of the strand exchange point, then the first 'A' which is itself on the 5' side of this 'T', must also be on the 5' side. Therefore I expect that the 1.9 kb band contains two labelled phosphates from the first 'A' and the 'T', and that the 1.2 kb band contains the phosphate from the second 'A'. Thus I reason from

the results of substrates 1 and 2, that the point of cleavage on the upper strand is between the last 'A' of the repeat element and the first 'T' of the spacer (figure 3.4b, substrate 2).

The results of digestion of the other two upper strand substrates, no.s 3 and 4, are also consistent with this interpretation. Substrate 3, labelled at two phosphates positioned either side of the deduced cleavage point, gives two bands on digestion of the recombination product (figure 3.4d lane 3). Substrate 4, labelled at a single phosphate positioned on the 3' side of the deduced cleavage point, gives a single band of 1.2 kb (fig 3.4c panel 4, lane C).

To determine the point of cleavage on the "lower" strand, I designed substrates 5, 6 and 7 (figure 3.4c). ScaI digests of recombination products are shown in fig 5e. Substrate 5, labelled at a single phosphate, positioned on the 5' side of the first 'T' in the lower strand of the spacer, gives a single band of 1.2 kb (figure 3.4e lane 5), consistent with the positioning of this phosphate on the 5' side of the strand exchange point. Figure 3.4e lane 6 shows that the two subsequent phosphates, (substrate labelled at "GG") are positioned on the 3' side of the cleavage point, (a single band of 1.9 kb is seen). Therefore the cleavage of the lower strand occurs between the last 'A' of the repeat element and the first 'T' of the spacer. The result from substrate 7, which is labelled at 4 consecutive phosphates, is consistent with this interpretation. Digestion of this recombination product gives two bands (figure 3.4e lane 7). The intensity of the 1.9 kb band is about three times higher than that of the 1.2 kb band, consistent with the fact that only one of the labelled phosphates, that on the 5´ side of the first 'T', is on the 5' side of the strand exchange point, whereas the other three phosphates are on the 3' side.

Taken together, these results show that the point of strand exchange in the Kw mediated recombination reaction is between the last 'A' of the repeat element and the first 'T' of the spacer in both strands of the substrate. Thus the spacer is 7 nucleotides long.

3.2.5 Kw mediates recombination in mammalian cells

The following experiment was carried out by P.- O. Angrand, and for this reason, the data is not shown here. A cotransfection assay in 293 cells was used to test the activity of Kw recombinase in mammalian cells. Transfection of cells with reporter constructs for the Kw and FLP recombinases, pSVpaK and pSVpaZ11 respectively did not give lacZ expression (less than 10^{-6} [lacZ+] cells in the transient expression experiments). In contrast, when pSVpaK was cotransfected with a Kw expression vector (pCMV-Kw1), about 1% of total cells showed *lacZ* expression by X-Gal staining. A similar range (between 5 to 10%) of *lacZ* expressing cells was obtained when pSVpaZ11 and a FLP expression vector (pOG44) were cotransfected in 293 cells. Control experiments showed that cotransfection of the Kw expression vector with a FLP recombinase reporter plasmid (pSVpaZ11) or cotransfection of a FLP expression vector with a Kw reporter plasmid (pSVpaK) did not give rise to *lacZ* expressing cells. These results demonstrate that the Kw recombinase is able to mediate recombination in mammalian cells in culture, and that the target site specificities of FLP and Kw are distinct in this assay.

3.3 Discussion

The results presented characterise a new recombinase, Kw, from the pKW1 plasmid of the yeast *Kluyveromyces waltii*. *In vitro* translated Kw recombinase specifically binds to a DNA target site consisting of two 18 base pair inverted repeat elements, and recombines this site in a conservative manner, carrying out both intra- and intermolecular recombination. Identification of the strand exchange point defines the spacer length as 7 base pairs. In an *in vitro* recombination assay, the recombinase works optimally at 30°C. *In vitro*, Kw is extremely sensitive to salt, working optimally at 30mM NaCl and 2mM $MgCl_2$. In addition, we have demonstrated that the Kw recombinase is functional in a mammalian cell culture system.

The characterisation of this novel member of the yeast subfamily of integrase class site specific recombinases enables comparison with other members of the family, and throws light on the extent of similarity among them. Two other members of the yeast family, FLP from *S. cerevisiae* and "R" from *Z. rouxii* (Araki et al., 1992) have so far been identified as recombinases, and FLP has been extensively characterised (Sadowski, 1995). Other members of the integrase family which have been well characterised are the λ integrase and the Cre recombinase from *E. coli* phage P1. (for review, see Stark et al, 1992).

The Kw DNA target site identified confirms the putative target site identified by Chen et al (1992a). The inverted repeat of two 18 base pair elements is sufficient as a target for efficient recombination *in vitro* and *in vivo*. The fact that the third repeat element found in pKW1 is not required, raises the question of its physiological role. Whether it has any effect on the efficiency of Kw recombination remains to be seen. It has also been shown for FLP that although all three repeat elements are bound by FLP (Andrews et al., 1985), the third repeat element is not required for recombination. It has no effect on recombination rates *in vitro* (Senecoff et al., 1985). Its presence can however, apparently enhance the rate of intermolecular recombination in *E.coli* (Jayaram, 1985). In the case of the 2μ like circle from *Z. rouxii*, (which encodes the R

recombinase), the recombination target contains an inverted repeat adjacent to four direct repeats. All six repeats are shown to be bound by R protein by DNAseI footprinting (Araki et al., 1992). Again, the inverted repeat element is sufficient for recombination (Serre et al., 1992), but the effect of the other four repeat elements has not been evaluated.

The Kw target site repeat element of 18 bp is longer than that of the FRT (13 bp) and the R target site (12 bp). It is unlikely that the entire sequence is involved in specific recognition and binding by Kw. Results from methylation protection and ethylation interference suggest that FLP makes specific contacts to only the 10 base pairs of the target site adjacent to the spacer (Bruckner and Cox, 1986). Mutation analysis of the FRT has shown that only point mutations at four positions within the 7 base pairs adjacent to the spacer have a negative effect on *in vitro* recombination (Senecoff et al., 1988). Further analysis of DNA -protein interactions in the Kw system and comparison with those already identified for FLP may throw some light on the elusive question of the basis of DNA binding specificity in these recombinases.

The strand cleavage point was determined by identifying the point of strand exchange. The length of the Kw target site spacer, 7 base pairs, falls within the 6 to 8 base pair range identified for other integrase family recombinases. The consistency of the spacer length in this family may reflect a requirement for a certain distance between cleavage points, or a certain phasing of those cleavage points with respect to the faces of the DNA helix, imposed by the structural constraints within the recombination synapse. Whether Kw recombinase follows the "FLP paradigm" (Chen et al., 1992b), in which the catalytic tyrosine of one recombinase subunit apparently cleaves a DNA strand which is bound by another subunit, remains to be seen. The R recombinase appears to follow this model (Yang and Jayaram, 1994), whereas the λ integrase cleaves the half site to which it is bound (for review, see Stark and Boocock, 1995).

The high salt sensitivity of Kw recombinase is in contrast to that reported for FLP, which tolerates a broad range of salt conditions, from 0 to 300 mM NaCl, in the presence of up to 20mM $MgCl_2$, with

optimum at 200 to 300 mM NaCl in the absence of $MgCl_2$. (Meyer-Leon et al., 1984, Babineau et al., 1985, Meyer-Leon et al., 1987). The R recombinase, on the other hand, is also reported to be salt sensitive (Lee et al., 1992). The temperature at which Kw shows optimum recombination *in vitro* is 30°C, which is consistent with its origins as a yeast enzyme. The loss of activity at higher temperatures appears to be due to protein instability. This effect has also been observed for FLP recombinase (Buchholz et al., 1996).

The characterisation of a novel site specific recombinase has implications not only for the further understanding of the family to which it belongs, but also for the field of genome manipulation technology. The site specific recombinases FLP, R and Cre have been extensively used in a variety of organisms to engineer specific DNA rearrangements at defined sites (Kilby et al., 1993). The characterisation of Kw recombinase widens the repertoire of tools available for this kind of experiment. We have shown in a cotransfection experiment in 293 cells, that the Kw recombinase can recombine its target site in a mammalian cell culture system, thus demonstrating its potential for *in vivo* genome manipulation. In this context, the preference of Kw recombinase for temperatures below 37°C may restrict the choice of organism in which it is used. The possibility that this effect is mainly due to protein instability (figure 3.2e) suggests that it may be overcome by expressing sufficient levels of the protein. A further useful detail of the Kw system is that the inverted repeat target site, like those of FLP, R and Cre, can be read in both orientations without encountering a stop codon, a feature which is necessary if the site is to be placed in an open reading frame.

CHAPTER 4
Kinetic analysis of FLP and Cre recombinases

4.1 Introduction

The integrase family site specific recombinases FLP, from the yeast *Saccharomyces cerevisiae,* and Cre, from bacteriophage P1, have been extensively used to direct DNA rearrangements in a wide range of organisms (for review, see Kilby et al., 1993, and Chapter 1, section 1.5). Much of the work to date has been concerned with establishing and optimising experiments in which one or other recombinase is used alone, and little has been done to compare the properties of the two recombinases in similar experiments. If there are differences in the behaviour of FLP and Cre, then such a comparative analysis might enable each recombinase to be chosen for the application to which it is better suited, and eventually, mutant proteins might be designed in which specific properties are altered for specific purposes.

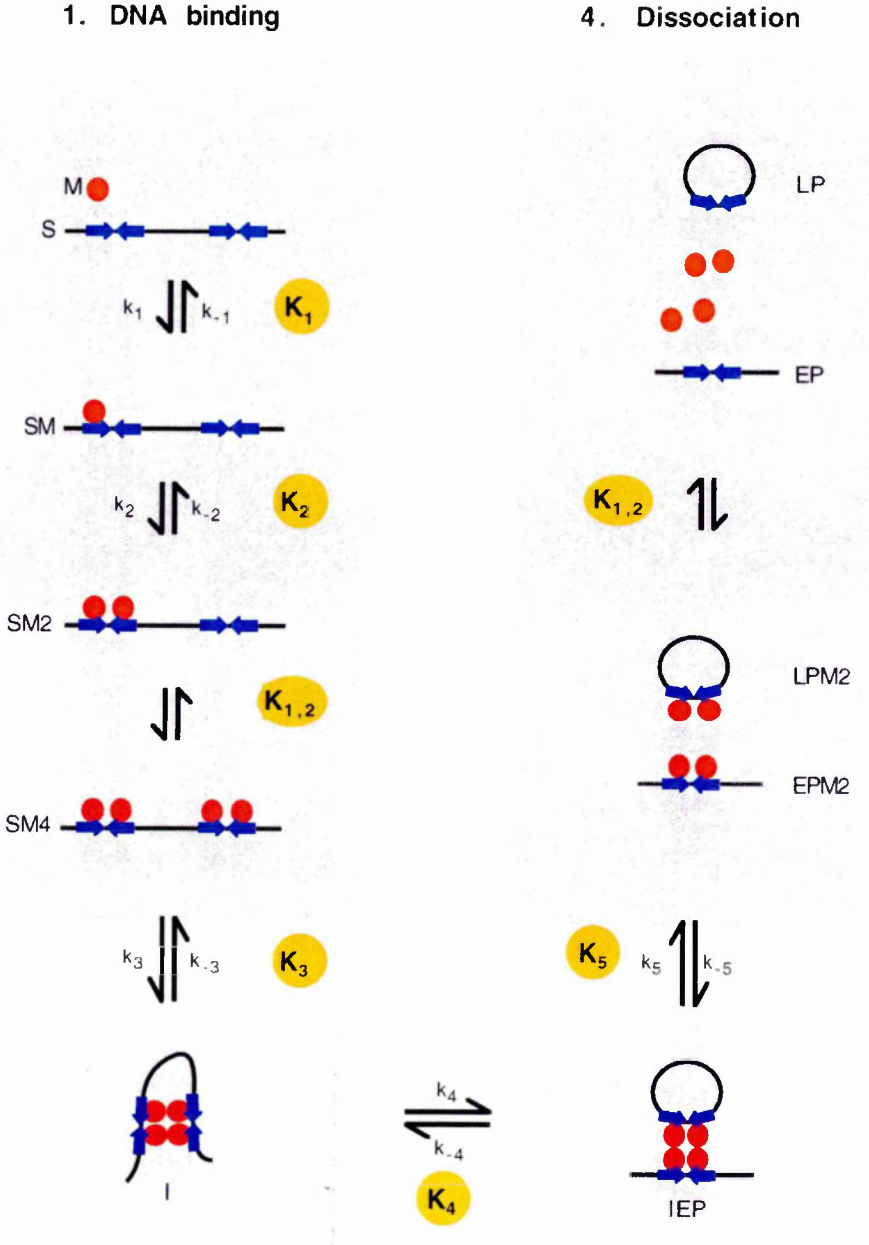
If a recombinase is used to recombine a chromosomally located target, then the enzyme has to find two copies of the target site in the entire genome. The specific affinity of the recombinase for its target is likely to determine the efficiency of this step. The non specific DNA binding affinity and the ability of the recombinase to recognise its target in chromatin, would also be expected to play a role. Once the target sites are bound, then the next step is synapsis, in which a single recombinase dimer bound to DNA has to find its partner. The efficiency of this process would be affected by the affinity of one recombinase dimer for another. The rate at which this process occurs would be expected to be particularly slow if the two target sites are on separate molecules, as is the case in chromosomal translocation or site specific integration.

As a first step towards determining and comparing some of these parameters for FLP and Cre, this chapter presents a kinetic analysis of the two recombinases based on *in vitro* studies. The steps in excisive recombination are shown in figure 4.1. Each step shown in the figure is reversible, and can be given a rate constant describing the backward and forward rate. Because the steps in the recombination reaction are interdependent, it is difficult to study them in isolation. Perhaps for this reason, little kinetic analysis of FLP and Cre recombination has been published.

The initial steps of DNA binding and dissociation can, however be studied using the gel mobility shift assay. Both FLP and Cre bind the inverted repeat target site first as a monomer, which is joined by a second monomer to form a dimer. (Andrews et al., 1987, Hoess and Abremski, 1984, Mack et al, 1992). Figure 4.1 shows that this process must occur twice to occupy both target sites of the excision substrate, and that the parameters governing the affinity of protein for recombined DNA, are the same as those which determine its affinity for the unrecombined substrate. Thus if the on- and off rates of each of the two recombinase monomers which consecutively bind the inverted repeat can be determined, then many steps of the reaction can be described. In the first part of this chapter, the DNA binding affinities of FLP and Cre are determined using the gel mobility shift assay, showing that Cre has a higher affinity for the LoxP site than FLP for the FRT, and that this is due both to a faster association rate, and to a slower dissociation rate, of both the first and second monomers.

From figure 4.1 it can be seen that if the DNA binding parameters are known, then the remaining processes in the recombination reaction are synapsis, catalysis, (which itself consists of multiple steps, see figure 1.3) and the dissociation of the synaptic complex after recombination. These processes are not easily dissected for kinetic analysis. In the second part of this chapter, the kinetics of the full excision reaction are compared for FLP and Cre, and the differences observed are described using a mathematical model. By introducing the measured DNA binding parameters for FLP and Cre into the model, the parameters for synapsis, recombination and synaptic dissociation which best describe the full recombination behaviour of each recombinase can be determined. This analysis suggests that the stability of the synaptic complex is higher for Cre than for FLP.

Figure 4.1 Steps in FLP and Cre excision recombination reaction . Inverted repeat elements are shown as blue arrows. Recombinase monomers are shown as red circles. Each step in the reaction is reversible and has a forward and a backward rate, indicated by black arrows. The forward and backward rate constants for each step are indicated on the figure (small k_n or k_{-n}). The equilibrium constant, (K_n) indicated on yellow circles, for the conversion of one complex into another is given by the quotient of the backward and forward rate constants. The name of each species as used for mathematical modelling is shown beside each complex.



4.2 Results I. DNA binding kinetics of FLP and Cre

4.2.1 Accuracy of the gel mobility shift assay: FLP and Cre complexes do not dissociate during electrophoresis.

The gel mobility shift assay is attractive for the analysis of stepwise binding of FLP and Cre to their target sites because it allows direct detection of distinct species. However the assumption that the assay accurately reports the distribution of different bound states is crucial to the kinetic analysis which follows. It has been suggested that the main source of error in the gel mobility shift assay is the dissociation of DNA -protein complexes during electrophoresis (Senear and Brenowitz, 1991, Hoopes et al., 1992).

It has been shown for the Gal repressor binding to a single operator site that less than 10% of protein- DNA complexes are lost by dissociation during electrophoresis (Senear and Brenowitz, (1991). For the yeast TFIID- TATA box complex on the other hand, considerable dissociation occurs (Hoopes et al, 1992). The extent of dissociation is dependent on the total time of electrophoresis, with half of the complexes being lost after 120 minutes. Clearly the question of whether or not dissociation occurs depends on the protein- DNA interaction being studied. To ask to what extent FLP and Cre protein -DNA complexes would survive electrophoresis, I used the methods of Senear and Brenowitz (1991) and of Hoopes et al. (1992).

The various DNA substrates used in the band shift assay are shown schematically in figure 4.2. Their preparation and quantification is described in "Materials and Methods". FLP or Cre protein was incubated at different concentrations with either a full site target, which contains two binding sites, or a mutated target (FRTM, LoxM) containing only one binding site (figures 4.3 and 4.4). At the range of concentrations used, FLP does not show non-specific binding to a non cognate target site (data not shown). At the highest Cre concentration (6.4 nM), a faint slowly migrating band is seen on both the LoxP and

Figure 4.2 Substrates used in the gel mobility shift assay.
(a) FRT full site target contains the inverted repeat from the wild type target. Inverted repeat elements are shown as *horizontal arrows*. A single mismatch at position 2 is indicated by *bold type*. Elements a and b are indicated above the sequence. The non palindromic spacer is indicated. Cleavage points are marked by *vertical arrows*. **(b) FRTM target.** In this substrate the b element of the FRT site is mutated to abolish binding. Positions which are mutated with respect to the wild type sequence are shown in *lower case*. **(c) FRT 1/2 site target.** The wild type target was cleaved at the XbaI site in the spacer, and the b element was isolated by gel purification. After end filling, 5 nucleotides of the spacer sequence remain. **(d) LoxP full site target** contains the inverted repeat from the wild type LoxP sequence. Inverted repeat elements are shown as *horizontal arrows*. Elements a and b are indicated above the sequence. The non palindromic spacer is indicated. Cleavage points are marked by *vertical arrows*. **(e) LoxM target.** In this substrate a SmaI site is introduced into the spacer, and the b element of the LoxP site is mutated to abolish binding. Positions which are mutated with respect to the wild type sequence are shown in *lower case*. **(f) LoxP 1/2 site target.** The loxM target was cleaved at the SmaI site in the spacer, and the a element was isolated by gel purification.

Figure 4.2

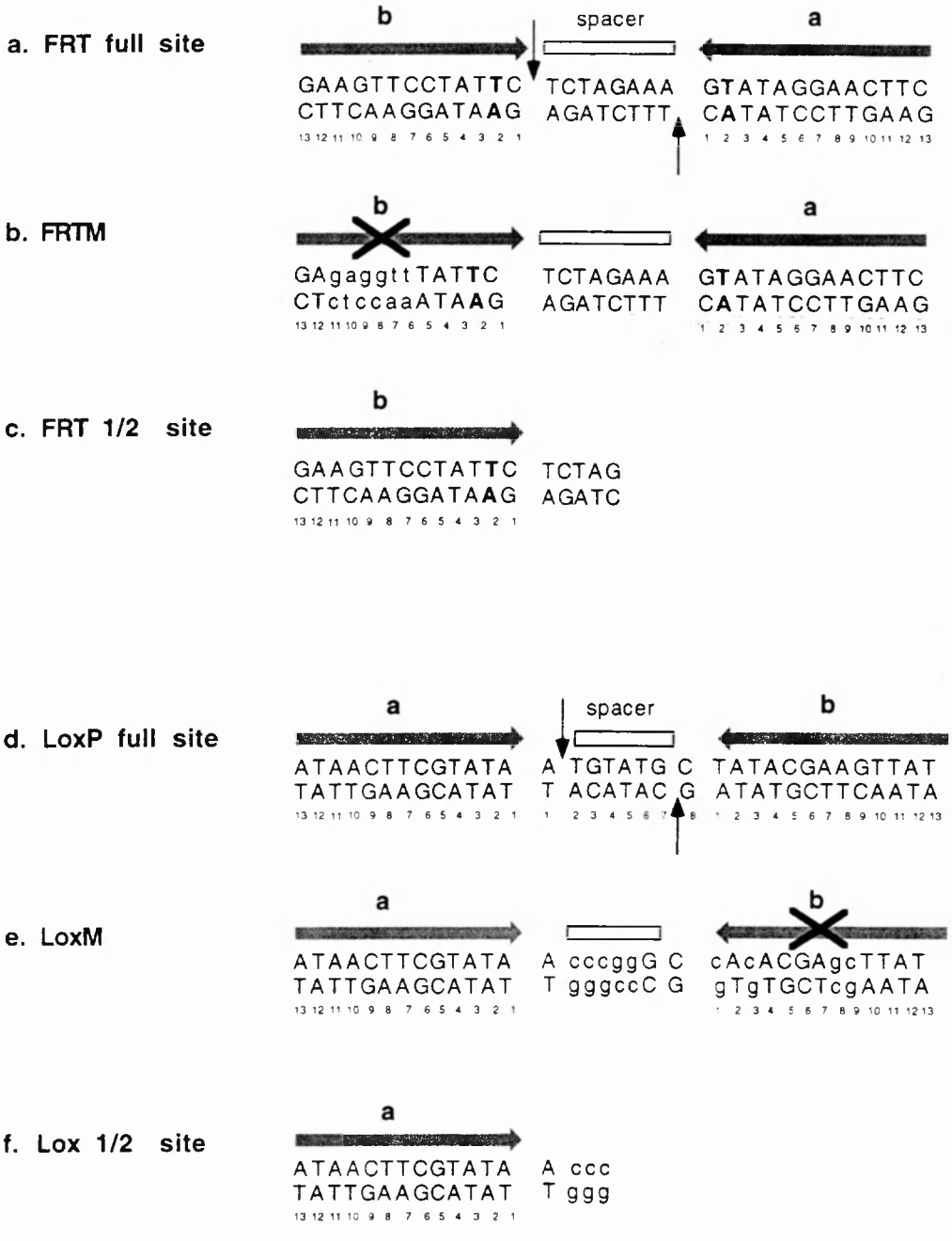

a. FRT full site
b
spacer
a
GAAGTTCCTATTC
CTTCAAGGATAAG
13 12 11 10 9 8 7 6 5 4 3 2 1
TCTAGAAA
AGATCTTT
GTATAGGAACTTC
CATATCCTTGAAG
1 2 3 4 5 6 7 8 9 10 11 12 13
b. FRTM
b
a
GAgaggttTATTC
CTctccaaATAAG
13 12 11 10 9 8 7 6 5 4 3 2 1
TCTAGAAA
AGATCTTT
GTATAGGAACTTC
CATATCCTTGAAG
1 2 3 4 5 6 7 8 9 10 11 12 13
c. FRT 1/2 site
b
GAAGTTCCTATTC
CTTCAAGGATAAG
13 12 11 10 9 8 7 6 5 4 3 2 1
TCTAG
AGATC
d. LoxP full site
a
spacer
b
ATAACTTCGTATA
TATTGAAGCATAT
13 12 11 10 9 8 7 6 5 4 3 2 1
A TGTATG C
T ACATAC G
1 2 3 4 5 6 7 8
TATACGAAGTTAT
ATATGCTTCAATA
1 2 3 4 5 6 7 8 9 10 11 12 13
e. LoxM
a
b
ATAACTTCGTATA
TATTGAAGCATAT
13 12 11 10 9 8 7 6 5 4 3 2 1
A cccggG C
T gggccC G
cAcACGAgcTTAT
gTgTGCTcgAATA
1 2 3 4 5 6 7 8 9 10 11 12 13
f. Lox 1/2 site
a
ATAACTTCGTATA
TATTGAAGCATAT
13 12 11 10 9 8 7 6 5 4 3 2 1
A ccc
T ggg

Figure 4.3 Gel mobility shift titration of FLP binding to (a) full site FRT target, and (b) mutant FRT target in which one site is inactivated. FLP was incubated for 10mins at 30°C at various protein concentrations as indicated with 0.2nM FRT(a) or FRTM (b). The nM FLP concentration is given above the figure; the log of the molar concentration is shown below the figure. Bands corresponding to free DNA, complexes containing one monomer (1) and 2 monomers (2) are marked.

(c) and (d) separate analysis of the protein- bound and free DNA bands from (a) and (b) respectively. (c) The sum of bands in complexes 1 and 2 in each lane of figure (a) was calculated as a fraction of total DNA in lane 1 (*closed triangles*, marked "bound). The free DNA was calculated as a fraction of the DNA in lane 1; this value was subtracted from 1 (*open circles*, marked "1-free"). Both values were plotted against the log of the molar concentration of free FLP monomer. (d) As for (c) except that the "bound" fraction represents complex 1 alone. The *solid lines* indicate the best fit curves obtained according to equation 3 as described in "Materials and Methods". Using the curves in figure 4.3d, values for the equilibrium association constant for FLP binding to a single site K_{1a} = 1.06 (±0.095) X 10^7 M^{-1} and K_{1a} = 1.87 (± 0.29) X 10^7 M^{-1} were obtained for the bound and free DNA bands, respectively.

Figure 4.3

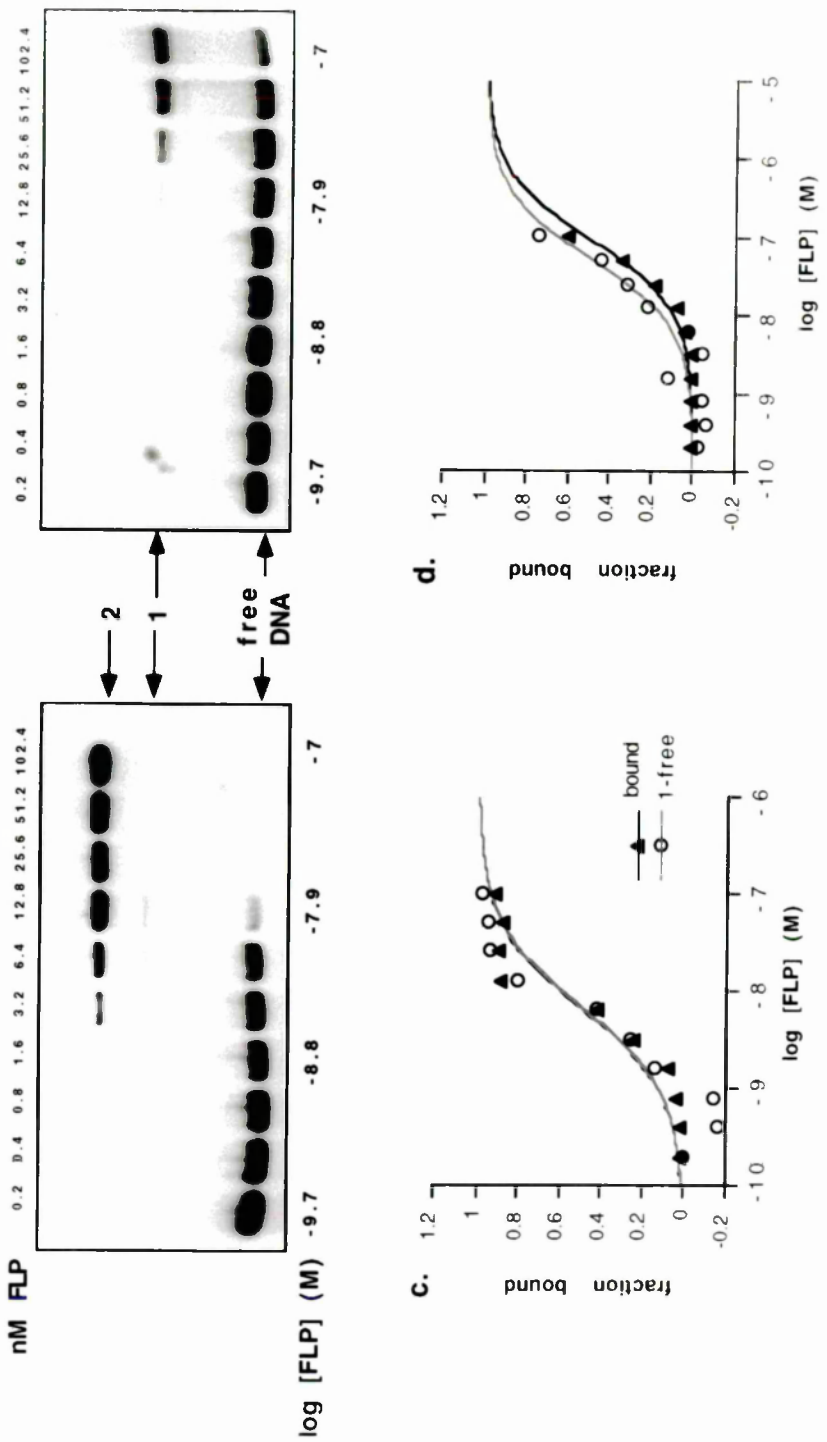

a. FRT 0.2nM
b. Mutant FRT 0.2nM
nM FLP
0.2 0.4 0.8 1.6 3.2 6.4 12.8 25.6 51.2 102.4
2
1
free DNA
log [FLP] (M)
-9.7 -8.8 -7.9 -7
c.
d.
fraction bound
bound
1-free
log [FLP] (M)

Figure 4.4 Gel mobility shift titration of Cre binding to (a) full site LoxP target, and (b) mutant LoxP target in which one site is inactivated. Cre was incubated for 10mins at 37°C at various protein concentrations as indicated with 0.2nM LoxP (a) or LoxM (b). The nM total input Cre concentration is given above the figure; the log of the molar input concentration is shown below the figure. Bands corresponding to free DNA, and to complexes containing one monomer (1) and 2 monomers (2) are marked.

(c) and (d) separate analysis of the protein- bound and free DNA bands from (a) and (b) respectively. (c) The sum of bands in complexes 1 and 2 in each lane of figure (a) was calculated as a fraction of total DNA in lane 1 (*closed triangles*, marked "bound). The free DNA was calculated as a fraction of the DNA in lane 1; this value was subtracted from 1 (*open circles*, marked "1-free"). Both values were plotted against the log of the molar concentration of free Cre monomer, calculated as described in "Materials and Methods". (d) As for (c) except that the "bound" fraction represents complex 1 alone. The *solid lines* indicate the best fit curves obtained according to equation 3 as described under "materials and methods". Using the curves in figure 4.4d, values for the equilibrium association constant for Cre binding to a single site K_{1a} = 1.67 (± 0.16) X 10^8 M^{-1} and K_{1a} = 2.55 (± 0.39) X 10^8 M^{-1} were obtained for the bound and free DNA bands, respectively.

Figure 4.4

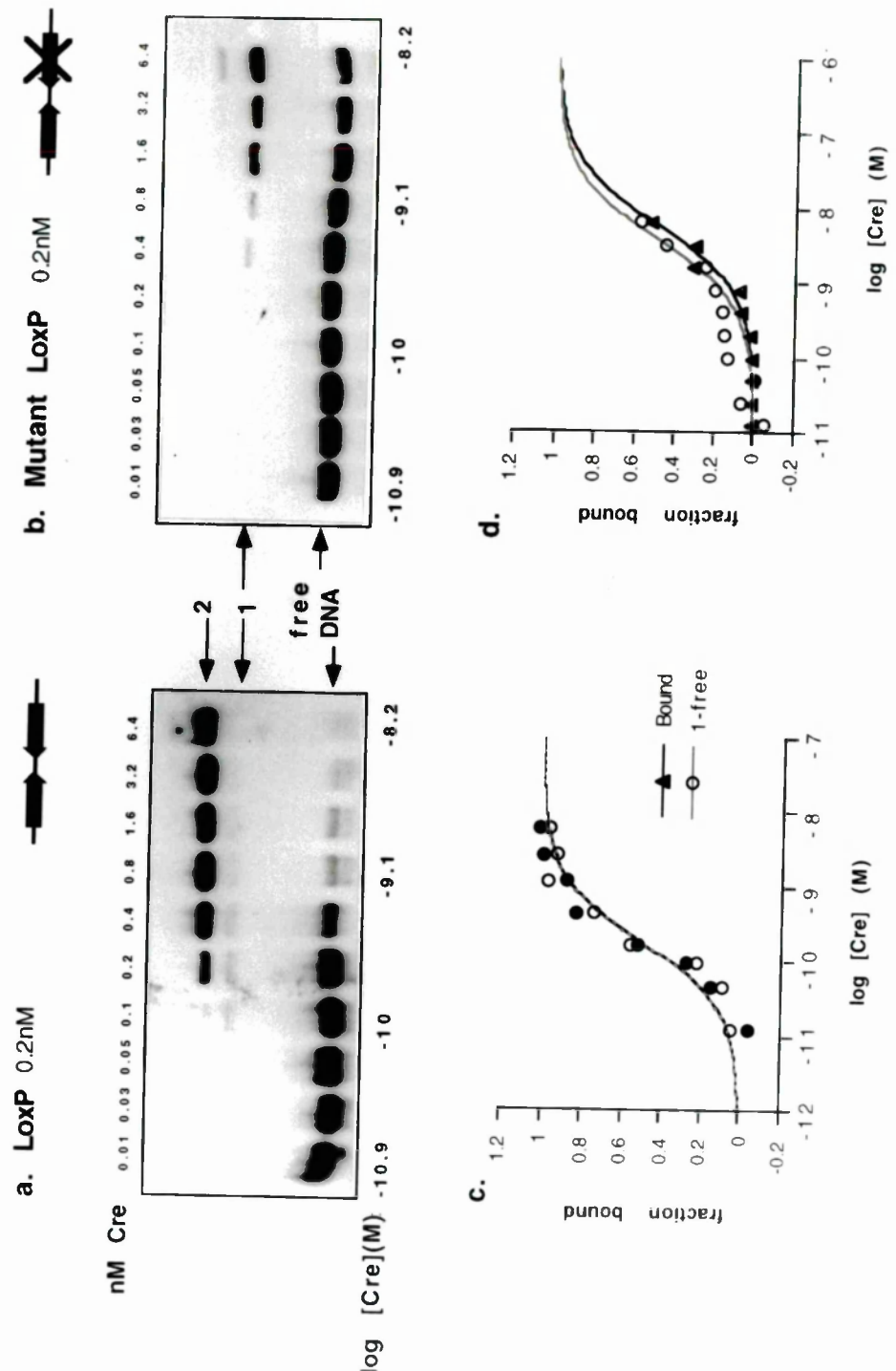
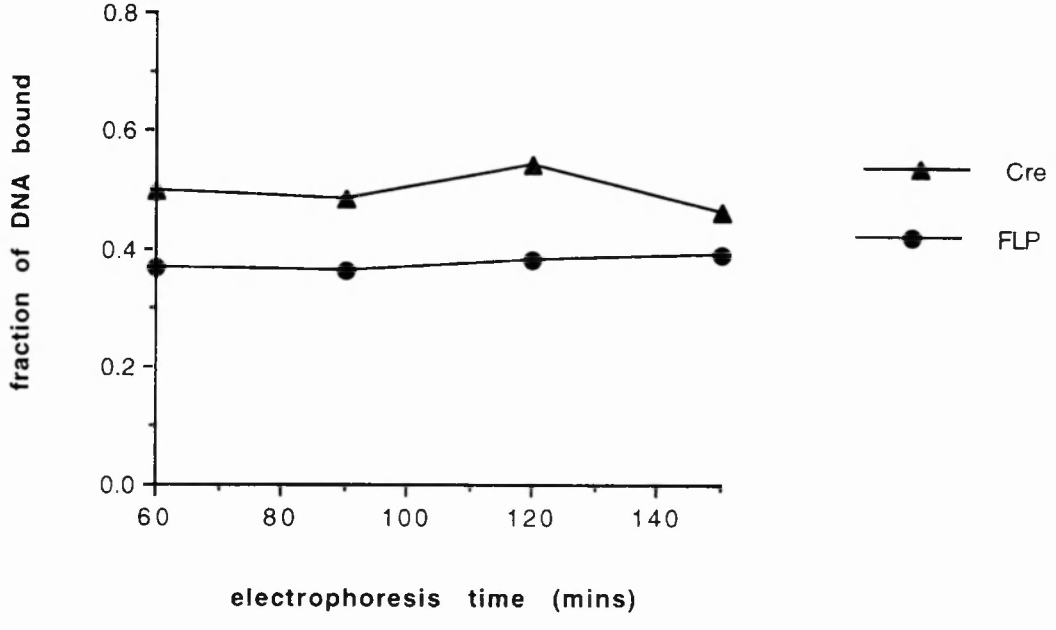

a. LoxP 0.2nM
b. Mutant LoxP 0.2nM
nM Cre
0.01 0.03 0.05 0.1 0.2 0.4 0.8 1.6 3.2 6.4
2
1
free DNA
log [Cre](M)
-10.9
-10
-9.1
-8.2
c.
d.
fraction bound
Bound
1-free
log [Cre] (M)

Figure 4.5 FLP and Cre complexes do not decay during electrophoresis. FLP and Cre binding reactions were carried out as described in "Materials and Methods". 6.4 nM FLP or 0.4 nM Cre were incubated with 0.2nM of the appropriate full site substrate. Reactions were started at 30 minute intervals, incubated for 10 mins and loaded immediately onto the gel, such that upon stopping the gel, complexes had been electrophoresed for 60, 90, 120 and 150 minutes. The fraction of DNA bound was calculated for each lane as (complex 1 + complex 2)/(complex 1 + complex 2 + free DNA).



the LoxM substrates (figure 4.4a, b, last lane, 6.4 nM Cre). This band may represent non specific binding, since the substrates used have about 15 base pairs on either side of the inverted repeat target, which could be bound non specifically by Cre. Alternatively, this band may represent a specific or non specific protein protein interaction. The identity of this band was not investigated further.

Figure 4.3a shows that FLP forms two complexes on the FRT full site (complexes 1 and 2). These complexes have been shown by others to correspond to occupation by a FLP monomer of one or both halves of the target site (Andrews et al., 1987, Lee and Jayaram, 1995b). On the mutated target site (figure 4.3b), only one complex is formed, which comigrates with the lower complex (complex 1), suggesting that it contains a monomer bound to one half of the FRT . Similarly for Cre, two complexes are formed on a full site target (complexes 1 and 2, figure 4.4a). Footprinting and stoichiometric analysis confirms that these are a monomer and a dimer complex respectively (Hoess and Abremski, 1984, Mack et al., 1992). On the mutated target LoxM, Cre forms a single specific complex which comigrates with complex 1 (figure 4.4a, b), suggesting that this represents a monomer bound to one half of the mutated target site.

The method of Senear and Brenowitz (1991) for detecting dissociation of complexes during electrophoresis is based on separate quantification of bound and unbound DNA. If complexes dissociate then some of the labelled DNA will form a smear ahead of the main protein DNA complex, whereas the free DNA band will be unaffected. Instead of treating each band as a fraction of the total in the lane, the bound and free DNA are analysed as individual bands. The bound and free DNA bands in figures 4.3 a, b and 4.4 a, b were quantified by PhosphorImager analysis and the fraction of DNA in the band relative to the DNA in lane 1 of each gel was determined (shown as "fraction bound" in the figure). The data were plotted against the concentration of free protein monomer, as shown in figures 4.3 c, d, and 4.4 c, d. The titration data is described by the Langmuir isotherm:

(fraction bound) = K.M/(1+K. M).

(Senear and Brenowitz, 1991). M is the concentration of free monomer. For free DNA,

(fraction bound)= (1- free DNA).

The data in figures 4.3 a, b and 4.4 a, b were fitted to the equation to generate the curves shown in figures 4.3 c, d and 4.4 c, d. These curves will be coincident only if dissociation does not occur. For total FLP binding, (figure 4.3c) and total Cre binding, (figure 4.4c) the fitted curves are almost exactly coincident, showing that there is no detectable loss of binding during electrophoresis. This result was confirmed for complex 2 using the method of Hoopes et al, (1992) in which complexes are electrophoresed over different time periods and the extent of binding compared. FLP and Cre dimer complexes do not detectably dissociate over electrophoresis times of up to two and a half hours (figure 4.5)).

The total DNA binding shown in figures 4.3c and 4.4c comprises both the monomer and the dimer complexes. Since the monomer complex represents such a small fraction of the total DNA bound for both FLP and Cre, the fact that there is no overall dissociation does not rule out the possibility that extensive dissociation of monomer complexes occurs but is not detected in this assay. For this reason the dissociation of monomer complexes was evaluated independently, using the data in figures 4.3c and 4.4c for the substrates FRTM and LoxM. Analysis of the fitted curves for both the FLP monomer complex (figure 4.3d) and the Cre monomer complex (figure 4.4d) indicates slightly tighter binding predicted by the curve for the free DNA than by the curve for the bound DNA, suggesting that some dissociation has occurred. For both FLP and Cre the fitted K values for the two curves are distinguishable (i.e the confidence limits do not overlap). The loss of monomer complexes represents less than 10% in both cases.

Table 4.1 Equilibrium association constants for FLP and Cre binding to DNA

K_{1a} is the equilibrium association constant for recombinase binding to a single element of a mutated target site and was calculated from the data in figures 4.3b and 4.4b as described in "Materials and Methods". K_1 and K_2 were calculated from figures 4.6 and 4.7 as described in "Materials and Methods". K_1 is the sum of equilibrium constants for recombinase binding separately to each site in a full site substrate ($K_1 = K_{1a} + K_{1b}$). K_{1b} was calculated as $K_1 - K_{1a}$. K_2 is the equilibrium constant for recombinase binding simultaneously to both sites of a full site target.

	FLP	$\Delta G_{(1)}$	Cre	$\Delta G_{(2)}$
K_{1a} (M^{-1})	1.06×10^7	9.87	1.67×10^8	11.82
K_{1b} (M^{-1})	1.08×10^7	9.88	4.33×10^8	12.41
K_1 (M^{-1})	2.14×10^7	10.30	6.00×10^8	12.61
K_2 (M^{-1})	8.92×10^8	12.57	7.4×10^{10}	15.62

(1) In kcal/mol at 30°C
(2) In kcal/mol at 37°C

The K value obtained from fitting the mutated site data gives the equilibrium association constant, K_{1a}, for a FLP or Cre monomer binding to a single binding site. (see Table 4.1). For FLP, the K value obtained from the free DNA data is 1.8 fold greater than that for the bound DNA, corresponding to a free energy difference of less than 0.35 kcal/mol. For Cre, the K value obtained from the free DNA data is 1.5 fold greater than that for the bound DNA, corresponding to a free energy difference of 0.2 kcal/mol.

In summary, these results show that for both FLP and Cre the dimer complex does not detectably dissociate during electrophoresis, whilst for the monomer complex, less than 10% dissociation occurs. I conclude that the gel mobility shift assay can provide data of sufficient accuracy for quantitative analysis of FLP and Cre.

4.2.2 Determination of the equilibrium association constants K_1 and K_2 shows that Cre has a higher DNA binding affinity than FLP.

The equilibrium constant K_{1a} determined above from mutated target sites (Table 4.1) indicates the affinity of a FLP or Cre monomer for a single site. In a full site system, complexes 1a and 1b (monomer binding to element a or b respectively) are not distinguishable and the macroscopic equilibrium constant K_1 for the observed complex 1, is equal to the sum of the microscopic constants for intrinsic binding to each site, $K_{1a} + K_{1b}$.

From the full site titration data in figures 4.3a and 4.4a, the equilibrium constants K_1 (complex 1) and K_2 (complex 2) were obtained using the method of Senear and Brenowitz (1991). The fractions of free DNA, complex 1 and complex 2 were plotted separately against free monomer concentration (figures 4.6 and 4.7). The data were fitted to equations 2a to 2c as described in "Materials and Methods", to yield values for K_1 and K_2 for FLP and Cre as shown in Table 4.1.

Figure 4.6 Determination of the equilibrium constants K_1 and K_2 from gel mobility shift titration of FLP bound to full site FRT. The fractions of DNA molecules with 0, 1 or 2 monomers bound were calculated from the bands in figure 4.3a as a fraction of the total DNA in each lane as described in "Materials and Methods". These values are plotted in figure 4.6 against log (molar concentration of free FLP monomer). *Open squares*: free DNA; *closed triangles*: FRT with 1 FLP monomer bound; *closed circles*: FRT with two monomers bound. The *solid curves* result from fitting the data according to equations 2a - 2c ("Materials and Methods"), to give $K_1 = 2.14 \times 10^7\ M^{-1}$ and $K_2 = 8.92 \times 10^8\ M^{-1}$

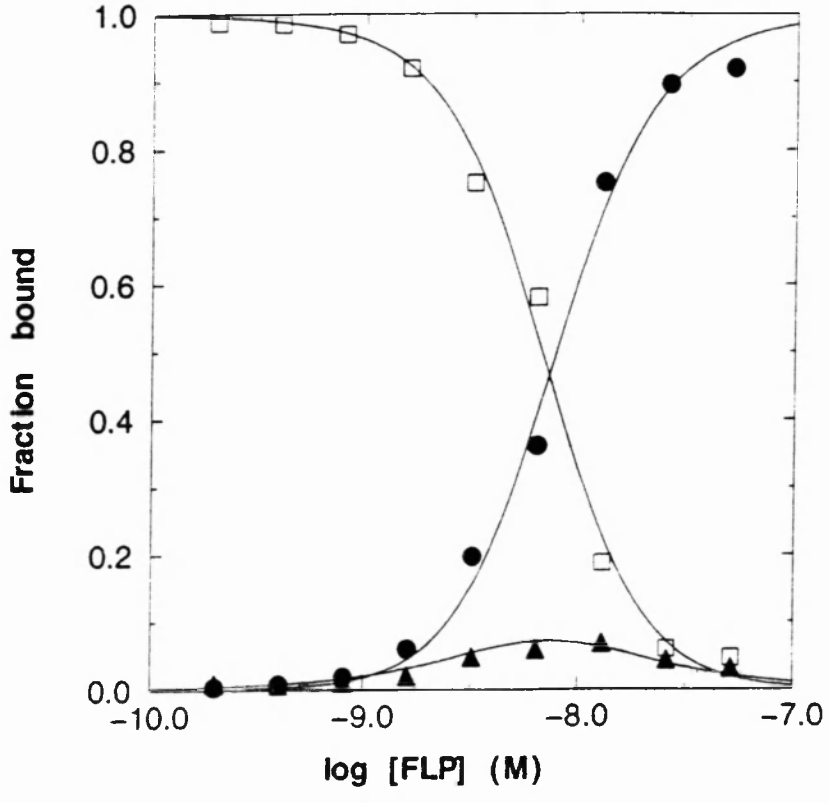
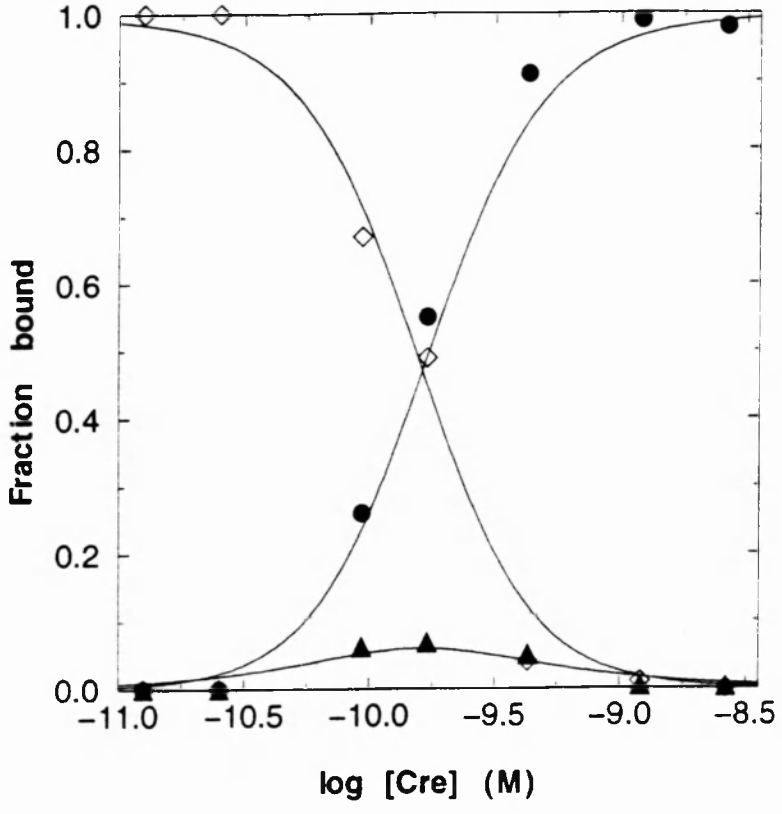


Figure 4.7 Determination of the equilibrium constants K_1 and K_2 from gel mobility shift titration of Cre bound to full site LoxP. The fractions of DNA molecules with 0, 1 or 2 monomers bound were calculated from the bands in figure 4.4a as a fraction of the total DNA in each lane as described in "Materials and Methods". These values are plotted in figure 4.7 against log (molar concentration of free Cre monomer), calculated as described in "Materials and Methods". *Open diamonds*: free DNA; *closed triangles*: LoxP with 1 Cre monomer bound; *closed circles*: LoxP with two monomers bound. The *solid curves* result from fitting the data according to equations 2a - 2c ("Materials and Methods"), to give $K_1 = 6.0 \times 10^8\ M^{-1}$ and $K_2 = 7.4 \times 10^{10}\ M^{-1}$.



Comparison of the K_2 values for FLP and Cre (Table 4.1: 8.92 X 10^8 and 7.4 X 10^{10} M^{-1} respectively) shows that the Cre dimer complex has a higher affinity for Lox P than the FLP dimer for the FRT. This corresponds to a free energy difference of 3.05 kcal/mol. The K_1 values for FLP and Cre (Table 4.1: 2.14 X 10^7 and 6.00 X 10^8 M^{-1} respectively) show that the monomer complex has a lower affinity than the dimer complex in both cases. The binding affinity of complex 1 for Cre is higher than that for FLP, corresponding to a free energy difference of 2.31 kcal/mol.

In order to dissect the constant K_1 into its components K_{1a} and K_{1b} it is necessary to compare the intrinsic affinities for separate sites. It has been shown for FLP by footprint analysis on a wild type target site containing an additional direct repeat element adjacent to the b element, that the b element is occupied before the a element (Andrews et al., 1987, Beatty and Sadowski, 1988; figure 4.2). This result suggests either that FLP has a higher intrinsic affinity for the b element, or that the presence of an additional repeat element next to the b element influences the order of site occupation. The FRT substrates used here lack the third repeat element. The mutant FRT site from which the constant K_{1a} was obtained contains only the a element. For FLP, comparison of the measured constant K_{1a} and the predicted constant K_{1b} (Table 4.1) does not suggest a significant difference in intrinsic binding affinities of FLP for the two halves of the target site. Separate determination of the constant K_{1b} using appropriate mutant target FRTs would be necessary to confirm this point. Consistent with this result, functional experiments in *E.coli* and a mammalian cell line with perfect a/a and b/b inverted repeats do not detect any difference in the recombination efficiencies of these targets, suggesting that there is no difference between the a and b half sites. (S. Chabanis, F. Buchholz and P.O. Angrand, unpublished results).

For Cre, the order of site occupation has not been studied. The target site repeat elements themselves are identical. If there is a difference in the intrinsic binding affinities of Cre for the two halves of its target site, then this may be expected to be mediated by contacts to the spacer region. The interaction of Cre with the

spacer is relatively well defined. DNAse footprinting analysis shows that the Cre dimer complex protects the entire spacer region as well as the inverted repeats (Hoess and Abremski, 1984). With neocarzinostatin, the same authors show that only one specific nucleotide in the spacer is not protected (the T on the top strand at position 6 in figure 4.2d). In addition, mutation analysis shows that a base change at position 5 of the spacer drastically reduces recombination even when the two recombining LoxP spacers are homologous (Hoess et al., 1986; figure 4.2d). Thus it is likely that Cre makes specific contact with at least one base in the spacer.

Comparison of the measured constant K_{1a} and the predicted constant K_{1b} for Cre (Table 4.1) shows that K_{1a} is 2.6 times smaller than K_{1b}, (a free energy difference of 0.59 kcal/mol). This result must however be interpreted with caution in view of the fact that in the LoxM target used in this assay, 5 nucleotides of the spacer are mutated, including the AT base pair at position 5 to which specific contacts are probably made (see figure 4.2e). It is possible that these changes reduce the intrinsic affinity of Cre for the single target site, and that the analogous experiment using the other half site (the b element) with similarly mutated spacer, would give a similarly low constant K_{1b}. Analysis with single site targets with wild type spacer sequences would be necessary to accurately determine K_{1a} and K_{1b}.

4.2.3 Both FLP and Cre bind cooperatively to a full site target.

Cooperatively interacting binding systems are characterised by decreased populations of intermediate configurations (Senear and Brenowitz 1991). From figures 4.6 and 4.7 it is clear that for both FLP and Cre the population of complex 1 is greatly reduced with respect to that of complex 2. Thus both systems show cooperative interactions of two bound recombinase monomers. This result is in agreement with the cooperativity already reported for FLP using footprint analysis (Andrews et al., 1987, Beatty and Sadowski, 1988). Cooperativity has not previously been demonstrated for Cre.

The difference between constants K_1 and K_2 gives an indication of the extent of cooperativity. For FLP, K_2 is about 40 fold greater than K_1. For Cre, K_2 is about 120 fold greater than K_1, indicating a stronger cooperativity for Cre than for FLP. In the light of this result it is interesting to note that FLP and Cre show differences in studies of spacer length variations, whose effect on recombination efficiency may reflect the importance of "cross core" interactions between the monomers bound at the two half sites. FLP tolerates spacer length changes, with some loss of recombination efficiency, of plus or minus one nucleotide provided both partner sites are homologous, (Senecoff et al., 1986) whereas Cre mediated recombination is abolished by addition of only one nucleotide to the spacer (Hoess et al., 1984).

4.2.4 Kinetic analysis of FLP and Cre DNA binding

The above analysis of FLP and Cre DNA binding behaviour yields information about DNA- protein interactions at equilibrium. For a full description of the kinetics of DNA binding, it is necessary to determine on and off rates for each species, and to see how their distribution evolves over time until equilibrium is reached.

Hoopes et al (1992) showed that the mobility shift assay can give accurate information for kinetic analysis of yeast TFIID binding to the TATA box. The authors monitored the binding of protein to its DNA target by loading aliquots of an ongoing reaction onto a bandshift gel over a time course. Association rate constants were determined from the initial binding rates. This system involves only one DNA- protein complex. However, its adaptation to a system in which two complexes are formed requires only that the two complexes can be unambiguously identified and that the bandshift assay accurately reflects their distribution relative to one another. Since these requirements are met for both FLP and Cre, I chose to use the bandshift assay for further kinetic experiments.

The substrates shown in figure 4.2 were used in the kinetic experiments. Full site substrates (FRT or LoxP) enable the binding of complex 1 and complex 2 to be observed simultaneously. As

described for the equilibrium experiments, complex 1 consists of both complexes 1a and 1b (recombinase monomer bound at element a or b respectively). To observe a recombinase monomer binding to a single element, the half site substrates were designed (see figure 4.2). The FRT half site contains element b. The LoxP half site contains element a. In addition the kinetics of binding to substrates FRTM, containing the a element of the FRT, and LoxM containing the a element of LoxP (figure 4.2) were also studied.

The kinetic experiments were carried out for FLP and Cre at four protein concentrations, chosen to cover the middle range of the equilibrium titrations shown in figures 4.3a and 4.4a. In addition, the highest concentration in each case (12.8nM and 1.6nM for FLP and Cre respectively) is that which is required for efficient recombination of 0.2nM DNA excision substrate *in vitro* (see chapter 4.3).

Figure 4.8 shows a time course of FLP at four concentrations binding to 0.4nM half site substrate (panel a) or 0.2nM full site substrate (panel b) Substrate concentrations were chosen so that the total nM half site concentration was constant. For the half site substrate (figure 4.8a, quantified in 4.8c), the initial rate of binding increases with protein concentration (compare the 15 second time point in each case). The maximum total binding reached also increases with protein concentration. However, it appears that a stable equilibrium is not reached within 5 minutes. Figure 4.8c shows that at all four FLP concentrations the amount of half site bound is slightly reduced at 5 minutes compared to 2 minutes. There are various possible explanations for this result, each of which were investigated as follows: The first possibility was that the reduction in volume of the reaction caused by removing aliquots at each time point led to aberrant behaviour at later time points, where the total volume was greatly reduced. Such behaviour may include for example, evaporation of liquid or increased aggregation of protein. This possibility was addressed by performing the experiment in a larger volume so that the reduction caused by removal of aliquots was negligible. The kinetics of FLP half site binding were identical to those observed in a smaller volume (data not shown).

Figure 4.8 Kinetics of FLP binding to full site and half site FRT measured by gel mobility shift assay. FLP protein was incubated with 0.4nM FRT half site (a) or 0.2nM FRT full site (b) at 30°C in 100µl volume at 4 different protein concentrations as indicated. 15µl aliquots were removed and loaded at the time points shown. Complexes containing one and two FLP monomers are marked beside the figure as 1 and 2 respectively. (c) and (d). The nM DNA in each of the bands in (a) and (b) was calculated as described in "Materials and Methods" and plotted against time for each experiment. (c). Half site. *Open triangles*: FRT half site bound by one FLP monomer. (d) Full site. *Closed triangles*: FRT site bound by one FLP monomer. *Closed circles*: FRT bound by two FLP monomers.

Figure 4.8

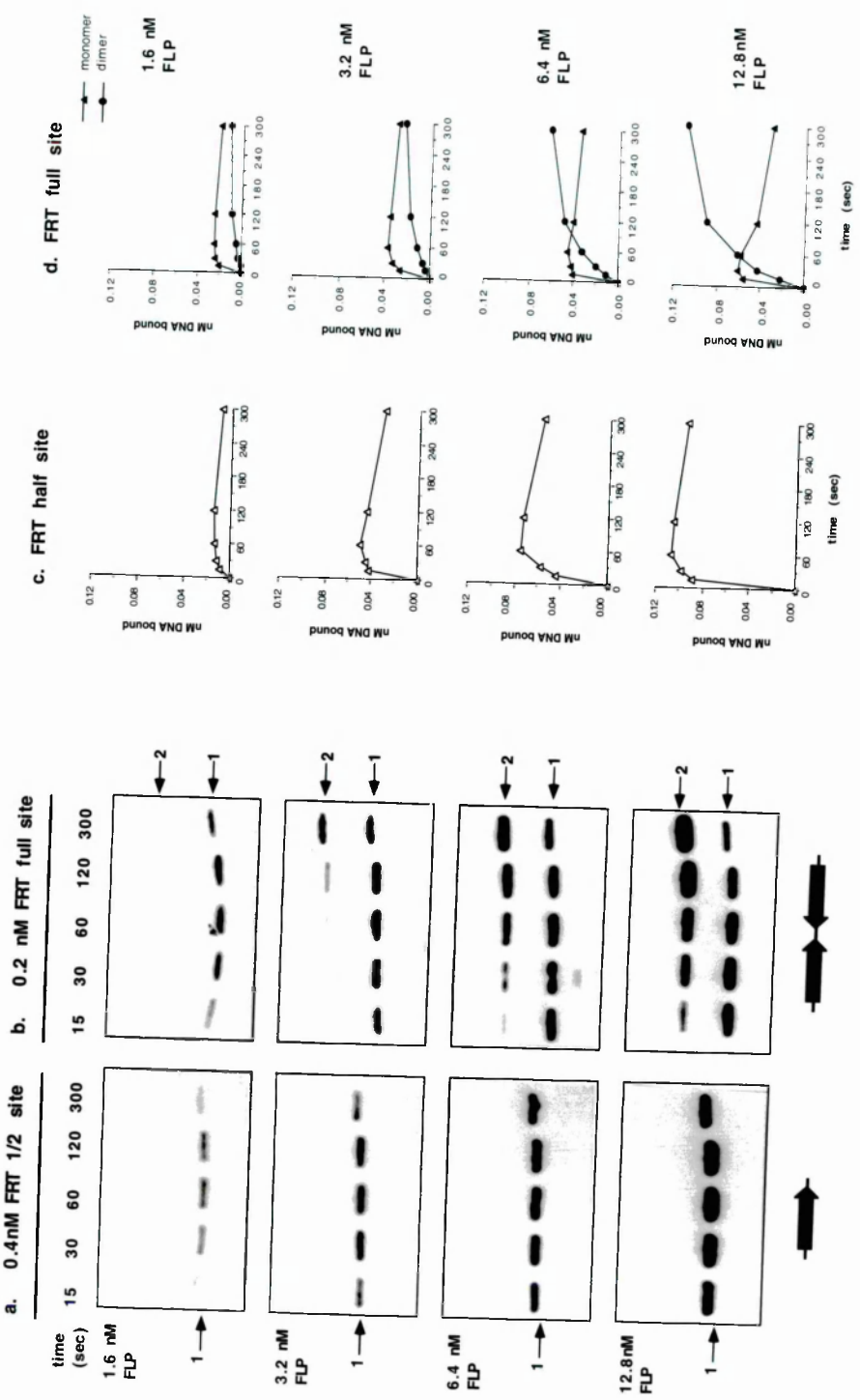
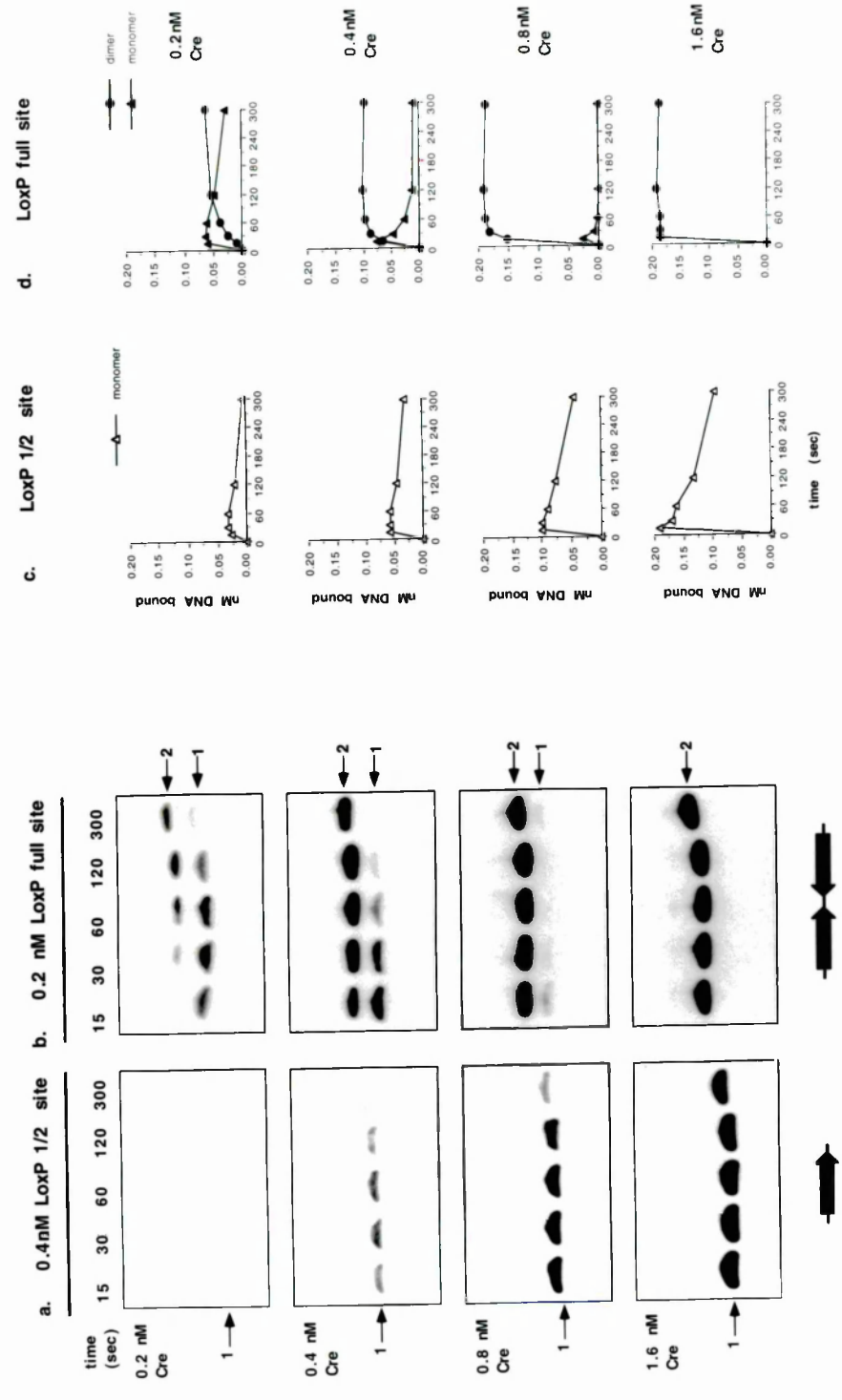

a. 0.4nM FRT 1/2 site
b. 0.2 nM FRT full site
time (sec)
15 30 60 120 300
1.6 nM FLP
3.2 nM FLP
6.4 nM FLP
12.8nM FLP
1
2
c. FRT half site
d. FRT full site
monomer
dimer
nM DNA bound
time (sec)

Figure 4.9 Kinetics of Cre binding to full site and half site LoxP measured by gel mobility shift assay. Cre protein was incubated with 0.4nM LoxP half site (a) or 0.2nM LoxP full site (b) at 37°C in 100µl volume at 4 different protein concentrations as indicated. 15µl aliquots were removed and loaded at the time points shown. Complexes containing one and two Cre monomers are marked beside the figure as 1 and 2 respectively. (c) and (d). The nM DNA in each of the bands in (a) and (b) was calculated as described in "Materials and Methods" and plotted against time for each experiment. (c). Half site. *Open triangles*: LoxP half site bound by one Cre monomer. (d) Full site. *Closed triangles*: LoxP site bound by one Cre monomer. *Closed circles*: LoxP bound by two Cre monomers.

Figure 4.9


a. 0.4nM LoxP 1/2 site
b. 0.2 nM LoxP full site
c. LoxP 1/2 site
d. LoxP full site
time (sec)
15 30 60 120 300
0.2 nM Cre
0.4 nM Cre
0.8 nM Cre
1.6 nM Cre
1
2
monomer
dimer
0.2nM Cre
0.4nM Cre
0.8nM Cre
1.6nM Cre
nM DNA bound
time (sec)

The second possible reason for the observed reduction of FLP binding over time was that the half site may act as a "suicide substrate". In cleaving its substrate, FLP becomes covalently linked to the 3´ phosphate at the point of cleavage, leaving a free 5´ OH group on the other side of the nick (Evans et al., 1990). This reaction is reversible, as long as the 5´OH group is present. In our assay the half site substrate is radioactively labelled at both ends, one of the labelled nucleotides being the C in top strand of the spacer (figure 4.2c). Cleavage of the phosphodiester bond adjacent to the spacer would result in the release of a pentanucleotide which may diffuse away, taking its radioactive label with it. Although such a process would trap the bound FLP monomer as a covalently linked molecule, the loss of label would mean that the total radioactivity in the observed band would decrease with each cleavage event. This possibility was investigated in two ways. Firstly, if a labelled pentanucleotide were released, then it would migrate ahead of the free DNA in the gel, and would not be included in the calculation of total DNA in the lane. If label were lost in this way, the sum of total DNA in each lane would decrease over time. However this was not the case. Instead the quantity of free DNA was observed to increase as bound DNA was reduced. Secondly, the experiment was repeated with the mutated full site substrate, FRTM, in which the cleaved strand cannot diffuse away. The kinetics of FLP binding to FRTM show a similar reduction in binding at later time points (data not shown). I conclude that the reduction in FLP binding to the half site over time may indicate an inherent characteristic of the protein.

Figure 4.8 panels b and d show the time course of FLP binding to a full site substrate. Both complexes 1 and 2 (monomer and dimer) are formed, and their changing distribution over time is observed. At the earliest time point, 15 seconds, the monomer is prevalent at all four protein concentrations. The dimer complex is formed subsequently, and the rate and extent of its formation increases with increasing protein concentration.

The monomer complex observed on a full site contains protein bound at site a or at site b. If the assumption is made that there is no significant difference in the affinity of FLP for sites a and b, then

twice as much monomer should be observed with the full site substrate as with the half site substrate. For the lowest concentration of FLP, 1.6nM, this is indeed the case (compare fig 4.8c and d, 1.6nM). At higher concentrations, the monomer complex is consumed increasingly rapidly by the dimer (figure 4.8d). On the full site substrate, the rate of monomer consumption at 3.2 nM, 6.4 nM and 12.8 nM FLP is faster than the rate of monomer complex loss observed with the half site substrate (figure 4.8c), showing that the conversion of monomer to dimer is faster than its inherent dissociation. With a full site substrate, both species reach a stable equilibrium between 5 and 10 minutes (not shown).

Figure 4.9 shows a time course of Cre at four concentrations binding to a half site substrate (panels a, c) or full site substrate (panels b, d). The initial rate of half site binding increases rapidly with protein concentration (compare 15 second time points in panel c). Half site binding is reduced at later time points. This reduction is more pronounced for Cre than for FLP. The same possible reasons for this result exist as discussed for FLP above, with the exception that for Cre, the half site spacer does not carry a radioactive label. The possibilities were tested in the same way as for FLP, and similar kinetics were observed. In the LoxM experiment (not shown), initial binding and subsequent reduction was similar to that observed for the half site. Note that the spacer mutations in the LoxP half site and LoxM substrates may be responsible for the accentuated loss of binding observed at later time points. To study the true kinetics of half site binding it would be necessary to design LoxP half sites and LoxM targets with wild type spacer sequences.

Figure 4.9 panels b and d show the time course of Cre binding to a full site substrate. At the lowest Cre concentration, 0.2nM, the monomer complex is prevalent at early time points, and is progressively replaced by the dimer complex. At higher Cre concentrations, (0.4 and 0.8nM), the monomer is consumed more rapidly by the dimer. At 1.6nM Cre, no monomer is detected at 15 seconds. The rate of monomer consumption at 0.4 nM, 0.8 nM and 1.6 nM Cre is faster than the rate of monomer complex loss observed at equivalent protein concentrations with the half site substrate

(figure 4.9c), showing that the conversion of monomer to dimer is faster than its inherent dissociation.

For Cre as for FLP, in the full site system, the observed monomer complex contains both complex 1a and 1b. As stated above, if Cre has equal affinity for elements a and b, then twice as much monomer should be observed with a full site substrate as with a half site substrate. At the lowest Cre concentration (0.2nM, figure 4.9d) the amount of monomer binding at 15 seconds (0.059 nM) is more than twice that observed for binding to the half site (0.026nM). Similar results were obtained with the LoxM target (not shown) consistent with the idea that Cre may have reduced affinity for the LoxM and LoxP half site substrates due to spacer mutations (see above). Again, experiments using substrates with a single Cre binding site and wild type spacer sequences would be necessary to resolve this point.

Comparing the behaviour of Cre to that of FLP on a full site target (figures 4.9d and 4.8d), it is clear that in this assay Cre shows a greater cooperativity of binding than FLP, indicated by the fact that for Cre, the monomer complex is replaced more rapidly and more completely by the dimer. This result is consistent with the equilibrium data described above. The fact that the same binding characteristics are observed in independent experiments underlines the faithfulness of the bandshift data for accurately reporting the behaviour of FLP and Cre.

4.2.5 Estimation of the association rate constants k_1 and k_2 shows that Cre has a faster DNA binding rate than FLP.

Using the initial binding rates observed in figures 4.8 and 4.9 and similar experiments, approximate values for the association rate constants were calculated as described in "Materials and Methods" (Table 4.2). The constants k_{1a} and k_{1b} for a FLP or Cre monomer binding to a single element were calculated from FRT and LoxP half site, FRTM and LoxM experiments as indicated in Table 4.2. The constant k_{1*} is the macroscopic constant for the sum of monomer binding to sites a and b, and was calculated from full site

experiments as described in "Materials and Methods". The constant k_2 for the binding of the second FLP or Cre monomer to a full site substrate was calculated from full site experiments. The values given in Table 4.2 are averaged over four protein concentrations, and the observed variation from this mean is given.

For FLP, the rate constants k_{1b} (half site) and k_{1a} (FRTM) do not differ significantly from one another (Table 4.2). The constant k_{1*} (full site) is approximately equal to the sum of k_{1a} and k_{1b} if the variation is taken into account. The association rate constant for the second monomer, k_2, is in the same range as k_{1a} and k_{1b} (Table 4.2).

For Cre, the values of all association rate constants are 20 to 100 fold greater than those measured for FLP (Table 4.2), showing that Cre binds its target site faster than FLP. The values for k_{1a} measured using either the LoxP half site or the LoxM site do not differ significantly from one another. Comparison of Cre k_{1a} with k_{1*} shows that k_{1*} is approximately 3 fold greater than $2(k_{1a})$. This is consistent with all other results using the LoxM or LoxP half site substrates and does not allow any conclusion to be reached regarding differential binding to the two halves of the site.

The apparent rate constant k_2 for the binding of the second Cre monomer increased with protein concentration, (from 1.3 to 3.2 X 10^8 M^{-1} s^{-1}.). This variation is very probably due to the fact that for all protein concentrations, the rate constant was measured using the data from the 15 second time point. To calculate k_2 , equation 13 ("Materials and Methods") was used:

$$k_2 = \frac{d[SM2]}{[SM][M].dt}$$

Because Cre binds its target very fast, the observed monomer (SM) and dimer (SM2) species at 15 seconds are misleading and do not give the true initial rate of binding. For example, at higher protein

Table 4.2 Estimated Rate constants for FLP and Cre binding to DNA

The estimated rate constants were derived from time course data as described in "Materials and Methods": k_1 (1/2 site) from the time course data for 1/2 site binding from figures 4.8a and 4.9a and similar experiments; k_1 (Msite) from time course experiments for binding to FRT M (FLP) or LoxM (Cre) substrates; k_{1*} and k_2 from the time course data for full site binding shown in figures 4.8b and 4.9b and similar experiments. The estimated dissociation constants k_{-1} and k_{-2} were calculated using equation 14 ("Materials and Methods") as follows: $k_{-1} = k_{1*}/ K_1$.; .$k_{-2} = k_2/ K_2$

	FLP	Cre
1) k_1 (M^{-1} s^{-1}) Association rate constant of first monomer		
k_{1b} (1/2 site)	1.8 (±0.6) X 10^6	3.4 (±0.9) X 10^7
k_{1a} (M site)	1.0 (±0.2) X 10^6	3.0 (±0.7) X 10^7
k_{1*} (full site)	3.8 (±0.9) X 10^6	2.0 (±1.3) X 10^8
2) k_2 (M^{-1} s^{-1}) Association rate constant of second monomer		
	2.8 (±0.5) X 10^6	1.3 to 3.2 X 10^8
3) k_{-1} (s^{-1}) Dissociation rate constant of first monomer		
	1.8 (±0.4) X 10^{-1}	3.3 (±2.1) X 10^{-1}
4) k_{-2} (s^{-1}) Dissociation rate constant of second monomer		
	3.1 (±0.6) X 10^{-3}	1.8 to 4.3 X 10^{-3}

concentrations (see e.g., 0.8nM Cre figure 4.9d) the observed concentration of SM at 15 seconds is low, and the concentration of SM2 is high. This would lead to an incorrect value for k_2 . To accurately measure k_2, earlier time points would have to be measured. (Note that for FLP the same reasoning holds, but there is less observed variation in k_2). The estimated dissociation rate constants k_{-1} and k_{-2} were calculated as described in Table 4.2 legend and "Materials and Methods". For both FLP and Cre the dissociation of the dimer (SM2) is calculated to be slower that that of the monomer (SM1), suggesting that DNA binding may be stabilised by the presence of a second monomer.

4.2.6 Mathematical simulation of FLP and Cre DNA binding.

The rate constants for FLP and Cre given in Table 4.2 are approximate values, measured from observed initial binding rates, and subject to the error inherent in taking samples at short time intervals. However, the true rate constants are responsible not only for the initial binding kinetics, but also for the behaviour of the monomer and dimer species over time, up to and after equilibrium. Therefore as a more rigourous means to find the true values for FLP and Cre rate constants, I have used a mathematical model in which the entire time course of DNA binding can be simulated and compared to the observed data. The model was constructed as described in "Materials and Methods", on the basis of the differential equations 7, 8, 15 and 16, which describe the evolution of species SM (monomer) and SM2 (dimer), M (free monomer) and S (free substrate) respectively, over time. The model represents a general description of a protein DNA interaction in which two linked sites are each bound by a protein monomer. The input values of k_1, k_{-1}, k_2 and k_{-2}, and of initial protein and substrate concentration can be varied to represent specific cases.

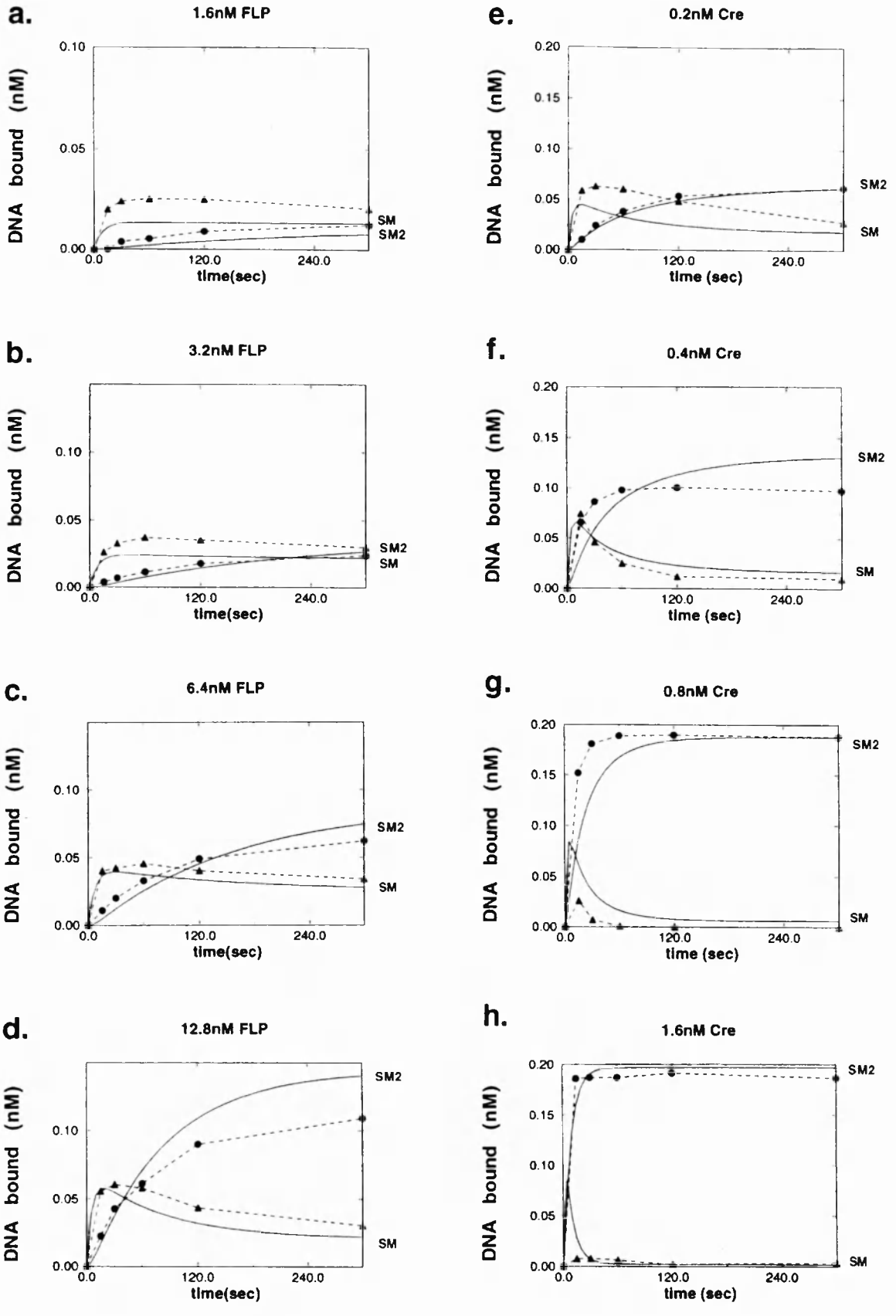
In figure 4.10, the model was used to simulate time course curves for DNA binding using the estimated rate constants for FLP and Cre (Table 4.2), at the protein and substrate concentrations used for the experiments in figures 4.8b and 4.9b. The input values for the rate constants k_{1*}, k_2 k_{-1} and k_{-2}Table 4.2) were chosen from within the

Figure 4.10 Simulation of time course curves using estimated parameters for FLP and Cre. The values for FLP and Cre for the measured rate constants k_1 and k_2 (see Table 4.2) and for the calculated rate constants k_{-1} and k_{-2} (see text), were used to simulate time course curves at different protein inputs. In each case values were chosen within the range of variation observed to give the best fit for all data sets. For FLP, the values used were: k_1=4.70 X10^6 $M^{-1}s^{-1}$; k_{-1}=1.00 X$10^{-1}s^{-1}$; k_2=2.10 X10^6 $M^{-1}s^{-1}$, and k_{-2}=3.73 X10^{-3} s^{-1}. The protein input values for FLP were (a) 1.6nM (b) 3.2nM (c) 6.4nM (d)12.8nM.

For Cre, the values used for the simulation were k_1=3.3 X10^8 $M^{-1}s^{-1}$; k_{-1}=1.30 X10^{-1} s^{-1}; k_2=1.30 X10^8 $M^{-1}s^{-1}$, and k_{-2}=1.80 X10^{-3} s^{-1}. The Cre protein input values were (e) 0.2nM (f) 0.4nM (g) 0.8nM (h) 1.6nM.

In each case, the measured data points from figures 4.8d and 4.9d are superimposed on the simulated curves. *Closed circles*: data points for complex 2; *closed triangles*: data points for complex 1. *Solid lines:* simulated time course curves. The curves corresponding to complexes 1 and 2 are marked SM and SM2 respectively.

Figure 4.10


a.
1.6nM FLP
DNA bound (nM)
0.10
0.05
0.00
0.0
120.0
240.0
time(sec)
SM
SM2
b.
3.2nM FLP
DNA bound (nM)
0.10
0.05
0.00
0.0
120.0
240.0
time(sec)
SM2
SM
c.
6.4nM FLP
DNA bound (nM)
0.10
0.05
0.00
0.0
120.0
240.0
time(sec)
SM2
SM
d.
12.8nM FLP
DNA bound (nM)
0.10
0.05
0.00
0.0
120.0
240.0
time(sec)
SM2
SM
e.
0.2nM Cre
DNA bound (nM)
0.20
0.15
0.10
0.05
0.00
0.0
120.0
240.0
time (sec)
SM2
SM
f.
0.4nM Cre
DNA bound (nM)
0.20
0.15
0.10
0.05
0.00
0.0
120.0
240.0
time (sec)
SM2
SM
g.
0.8nM Cre
DNA bound (nM)
0.20
0.15
0.10
0.05
0.00
0.0
120.0
240.0
time (sec)
SM2
SM
h.
1.6nM Cre
DNA bound (nM)
0.20
0.15
0.10
0.05
0.00
0.0
120.0
240.0
time (sec)
SM2
SM

range of variation observed, to give the best overall fit to all four data sets for each of FLP and Cre.

Comparison of the simulated curves for FLP (figure 4.10a) and Cre (figure 4.10b) shows that the model qualitatively reproduces the behaviour of FLP and Cre, even with approximate input values for the four parameters. For example the highly cooperative behaviour of Cre is reproduced by the model, with higher protein input values giving very rapid consumption of the species SM by the species SM2 (figure 4.10b, 0.8 and 1.6 nM). However, the fit of the simulated curves to the data is approximate in all cases. For FLP for instance, (figure 4.10a) the concentration of species SM (monomer) simulated by the model is consistently lower than that observed, suggesting that the measured value of k_1* may be an underestimate. For FLP, the simulated species SM2 fits the data reasonably well at 3.2 and 6.4 nM protein, but rises above the observed dimer concentration at 12.8 nM protein. For Cre, the concentration of the simulated species SM is below that of the observed monomer at 0.2 nM Cre input, fits the data relatively well at 0.4 nM Cre, but rises above the observed concentration at 0.8 and 1.6 nM Cre. (Note, however, that the peak in SM observed in the simulated curve arises before the 15 second time point, where the first real data point is taken). The simulated species SM2 for Cre fits the data best at 0.2 nM and 1.6 nM protein, and deviates from the measured dimer concentration at 0.4 and 0.8 nM. The fact that the deviation of the simulated curves from the data is not of a simple nature, is evidence that the binding kinetics of the species SM and SM2 are intricately coupled (this interdependence is described by the coupled equations 7 and 8).

It should be emphasised at this point, that the differences between the simulated curves and the measured data far exceed the experimental variation observed from repeated time course experiments. Experiments with identical protein input, mixed from the same protein and buffer mixtures show variations of 5% or less. The experiments in figures 7 and 8 were mixed from the same buffer mixture, using serial dilutions of the same protein aliquot, and thus variations of the same order should be expected between individual time course experiments. Within a single time course experiment, aliquots were taken from a single ongoing reaction, so

the principal source of error is the precision of timing. In addition it should be kept in mind that the relative distributions of the monomer and dimer species in each lane are faithfully reported by the bandshift assay. This means that the fit of the two simulated curves relative to each other is an important criterion.

4.2.7 Optimisation of the parameters k_1, k_{-1}, k_2 and k_{-2}: Four parameters are sufficient to describe FLP and Cre DNA binding.

To find the combination of parameters k_1, k_{-1}, k_2 and k_{-2} which best describes the DNA binding behaviour of FLP and Cre, the four parameters were varied simultaneously for each protein concentration individually. Figure 4.11 shows the fitted curves after optimisation of parameters. The parameters used to give each fit are shown in Table 4.3.

After the optimisation of parameters, a good fit was obtained for all time course experiments for both proteins, with the exception of the data set for Cre at 0.4 nM (figure 4.11f). This result shows that, in most cases, four parameters are sufficient to describe the DNA binding behaviour of FLP and Cre. The four rate constants were optimised for each time course experiment separately. Comparison of the optimised values for a given protein at different concentrations shows that the values are consistent with one another (Table 4.3). For example, both k_{-1} and k_{-2} for FLP are almost identical for all four protein concentrations. Slight variations in k_1 and k_2 (up to two fold) are seen, but the ratio of k_1 to k_2 is constant. The observed variation in the values of k_1 and k_2 can mainly be accounted for by the expected variation of 5% between individual experiments. For Cre, the values of k_{-1} and k_{-2} are identical at all protein concentrations, with the exception of 0.4 nM (Table 4.3). The values of k_1 and k_2 show variations of up to two fold, again with the exception of the experiment at 0.4 nM.

The fact that the model could not be fitted well for the Cre data at 0.4 nM protein input (figure 4.11b) and that the values of the four parameters used for the best fit obtainable vary greatly from those for the other Cre concentrations (Table 4.3), raised the possibility

Table 4.3 Optimised rate constants for FLP and Cre binding to DNA

The values of k1, k-1, k2 and k-2 resulting from the optimisation of curve fitting for FLP and Cre at each protein input (see text) are shown.
(1) A decay constant, describing inactivation of protein during the binding reaction was used to fit the 0.4nM Cre data set.
(2) 0.4 nM Cre data fitted without decay constant
(3) 0.4 nM Cre data fitted with decay constant

Protein input		Rate constant (k_n: M^{-1} s^{-1}. k_{-n}: s^{-1})				decay (1) (M^{-1} s^{-1})
		k_1	k_{-1}	k_2	k_{-2}	
1) FLP	1.6 nM	$1x10^7$	$1x10^{-1}$	$4x10^6$	$1.3x10^{-2}$	_
	3.2 nM	$7x10^6$	$1x10^{-1}$	$3x10^6$	$1.3x10^{-2}$	_
	6.4 nM	$6x10^6$	$1x10^{-1}$	$3x10^6$	$1.3x10^{-2}$	_
	12.8 nM	$4x10^6$	$7x10^{-2}$	$2.3x10^6$	$1.3x10^{-2}$	_
2) Cre	0.2 nM	$2.2x10^8$	$2.4x10^{-2}$	$1.2x10^8$	$1.8x10^{-3}$	_
	0.4 nM(2)	$5.0 x10^8$	$3.0x10^{-1}$	$2.5x10^8$	$8.0 x10^{-3}$	_
	0.4 nM(3)	$3.5x10^8$	$2.4x10^{-2}$	$3.0x10^8$	$1.8x10^{-3}$	$1.7x10^7$
	0.8 nM	$2.2x10^8$	$2.4x10^{-2}$	$3.0x10^8$	$1.8x10^{-3}$	_
	1.6 nM	$2.2x10^8$	$2.4x10^{-2}$	$3.0x10^8$	$1.8x10^{-3}$	_

Figure 4.11 Optimisation of parameters for FLP and Cre

The curves for SM and SM2 shown in figure 4.13 were fit to data at each protein input individually by simultaneous optimisation of the rate constants k_1, k_{-1}, k_2 and k_2. The optimised values for the four parameters are shown in Table 4.3. In each case, the data points from figures 4.8d and 4.9d are superimposed on the plot: *closed triangles*: complex 1; *closed circles:* complex 2. The simulated curves are shown by *solid lines*, marked SM (complex1) and SM2 (complex 2). (a) FLP, 1.6 nM. (b) FLP, 3.2 nM. (c) FLP, 6.4 nM. (d) FLP, 12.8 nM. (e) Cre, 0.2 nM. (f) Cre, 0.4 nM (g) Cre, 0.8 nM. (h) Cre, 1.6 nM.

Figure 4.11

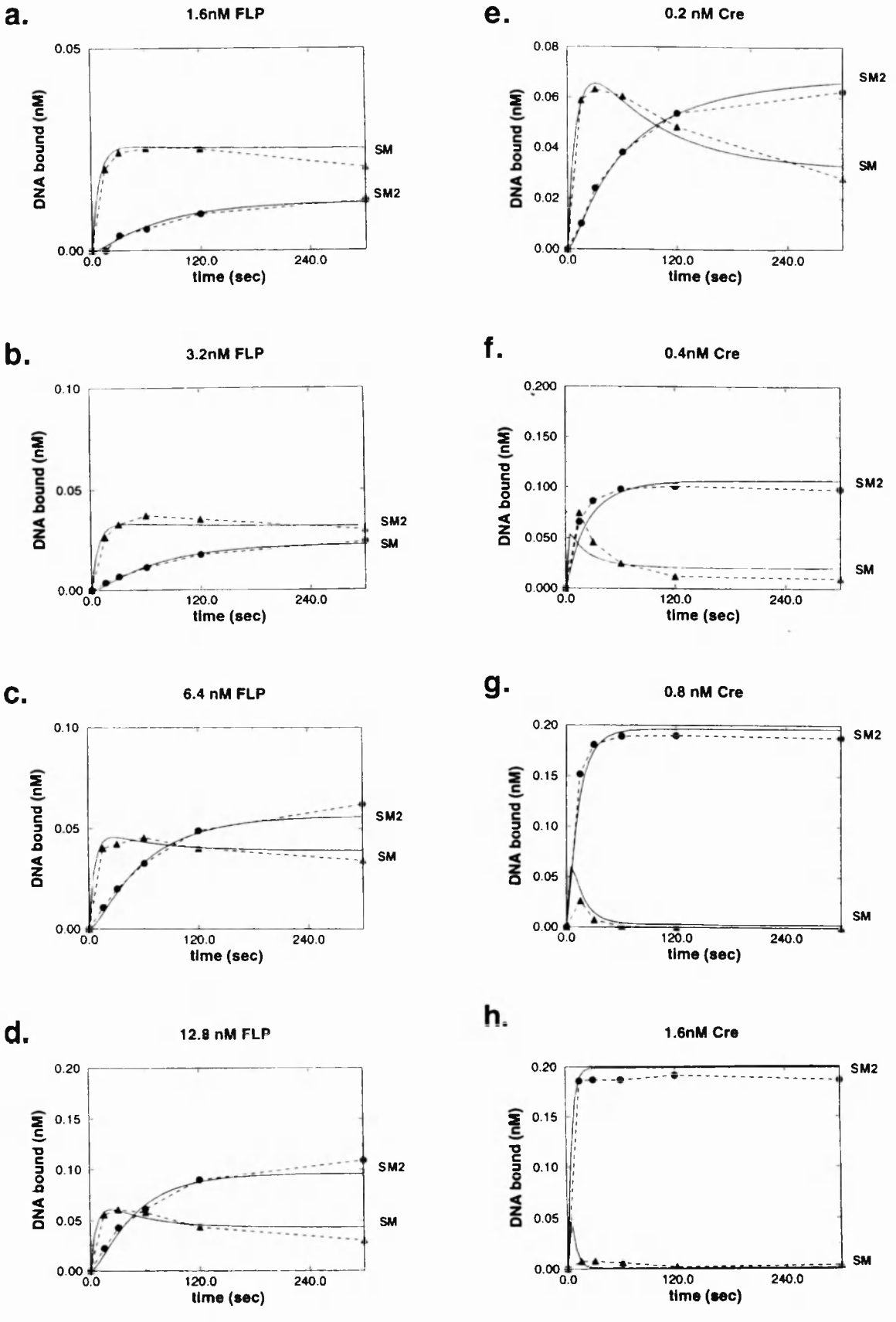
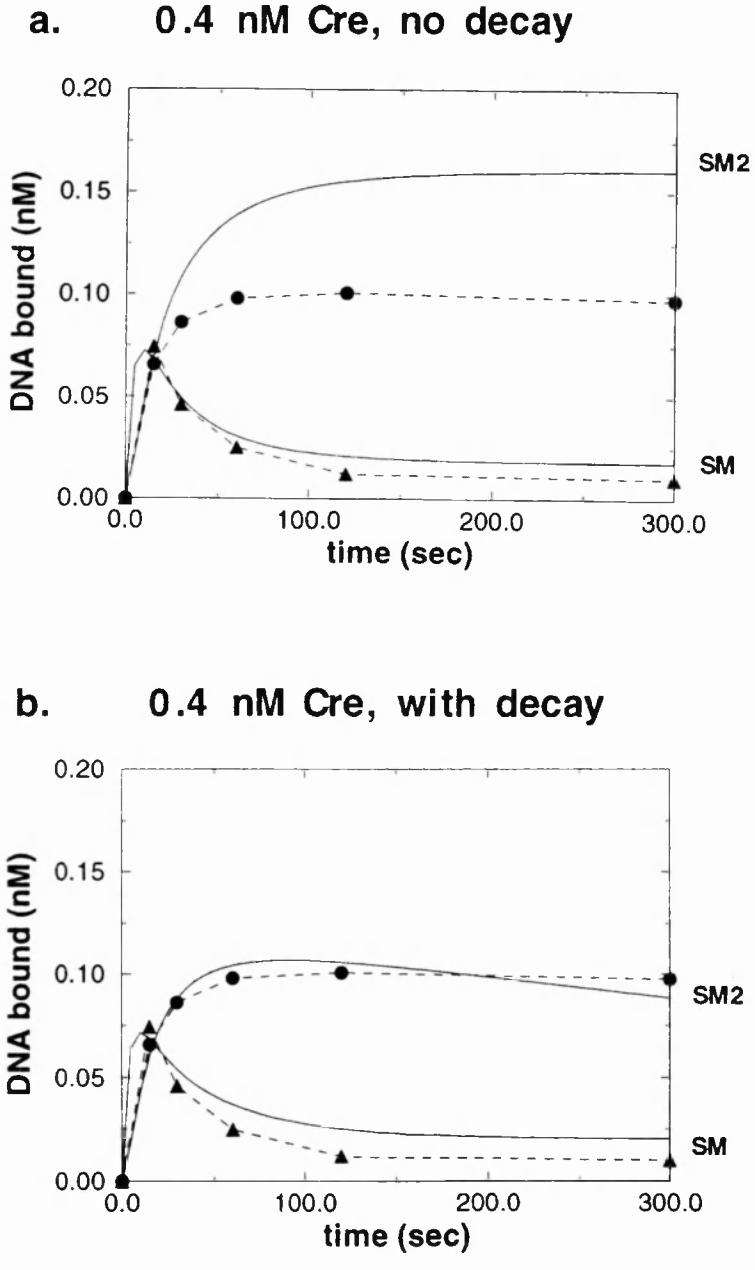

a.
1.6nM FLP
DNA bound (nM)
0.05
0.00
0.0
120.0
240.0
time (sec)
SM
SM2
b.
3.2nM FLP
DNA bound (nM)
0.10
0.05
0.00
0.0
120.0
240.0
time (sec)
SM2
SM
c.
6.4 nM FLP
DNA bound (nM)
0.10
0.05
0.00
0.0
120.0
240.0
time (sec)
SM2
SM
d.
12.8 nM FLP
DNA bound (nM)
0.20
0.15
0.10
0.05
0.00
0.0
120.0
240.0
time (sec)
SM2
SM
e.
0.2 nM Cre
DNA bound (nM)
0.08
0.06
0.04
0.02
0.00
0.0
120.0
240.0
time (sec)
SM2
SM
f.
0.4nM Cre
DNA bound (nM)
0.200
0.150
0.100
0.050
0.000
0.0
120.0
240.0
time (sec)
SM2
SM
g.
0.8 nM Cre
DNA bound (nM)
0.20
0.15
0.10
0.05
0.00
0.0
120.0
240.0
time (sec)
SM2
SM
h.
1.6nM Cre
DNA bound (nM)
0.20
0.15
0.10
0.05
0.00
0.0
120.0
240.0
time (sec)
SM2
SM

Figure 4.12 Effect of protein decay.

Simulated curves for SM and SM2 were fit to the data set for Cre at 0.4nM, (figure 4.9d) using only k_1, k_{-1}, k_2 and k_{-2} (a) or using the same values of these four parameters plus a decay constant (b), describing the loss of active monomers from the system during the binding reaction (see text for details). The data points from figure 4.9d, (Cre 0.4nM) are superimposed on each plot: *closed triangles*, complex 1; *closed circles*, complex 2. *Solid lines* , marked SM (complex 1) and SM2 (complex 2) indicate the simulated binding curves. The input values used for the simulation were **(a)** k_1= 3.5 X 10^8 M^{-1} s^{-1}; k_{-1}= 2.4 X 10^{-2} s^{-1}; k_2= 3.0 X 10^8 M^{-1} s^{-1}; k_{-2}= 1.8 X 10^{-3} s^{-1}. **(b)** k_1, k_{-1}, k_2, k_{-2}, as for (a). Decay constant =1.7 X 10^7 M^{-1} s^{-1}.



that in this experiment, aberrant behaviour of Cre may lead to an increased error. Such "aberrant behaviour" may be for instance, an inactivation of protein during the course of the experiment. Such inactivation may be a true decay of active protein, but may also be caused for example by aggregation of Cre into a complex which effectively removes monomers from the pool available, or by protein sticking to the sides of the tube during the reaction. The hypothesis that a "decay" factor plays a role in defining the observed binding kinetics is easily incorporated into the model.

Figure 4.12 shows the result of including a decay factor on the fitting of the model for Cre at 0.4 nM. Without a decay factor, using values for k_1, k_{-1}, k_2 and k_{-2} close to those for the other Cre protein concentrations, (Table 4.3), the fit shown in figure 4.12a is obtained. The species SM fits the observed monomer binding curve reasonably well, whilst the species SM2 overshoots the observed dimer binding curve by about 1.5 fold. In figure 4.12b, the same four rate constants are used, with the inclusion of a fifth parameter, describing decay of active monomers. The inclusion of the decay constant greatly improves the fit of the model: the species SM is virtually unaffected, whilst the species SM2 is reduced to the level of the observed dimer binding curve. This result shows that when an additional decay constant is included in the model, the observed data for Cre at 0.4 nM can be described by four rate constants similar to those for the other Cre concentrations. It also possible that the slight variations in the optimised rate constants for FLP and Cre (Table 4.3) are caused by a similar decay process (which may vary in individual experiments or be protein concentration dependent). The inclusion of a small decay parameter in the model for the other data sets may enable the model to be fitted to the data with identical binding rate constants for each protein concentration. However, since the variation in the optimised values of k_1 and k_2 for FLP and Cre is small this idea is not pursued further here.

4.2.8 Comparison of optimised rate constants: Models for FLP and Cre DNA binding.

The constant k_1 in Table 4.3 is the macroscopic association rate constant for monomer binding to site a or b of the FRT or LoxP target site, and is equal to $k_{1a} + k_{1b}$. If the binding rates for site a and b are identical, the microscopic rate constant for the binding of the first monomer to either side of the target site is given by $k_1/2$.

Comparison of the association rates shown in Table 4.3 for the binding of the first and second FLP monomers ($k_1/2$ and k_2 respectively), suggests a mechanism for the binding of the second monomer. Target site recognition by the second FLP monomer could be facilitated (or hindered) by the presence of a FLP monomer already bound. However, for FLP the constant k_2 is very nearly equal to $k_1/2$. This indicates that the second FLP monomer binds its DNA target site at much the same rate as the first, and suggests that the role played by protein-protein interactions in the recognition process is minimal. The cooperativity of binding of the second FLP monomer is therefore mediated mainly by a slow dissociation rate once both monomers are bound. The dissociation rate constant k_{-2} is about 10 fold smaller than k_{-1}.

For Cre, comparison of the values for $k_1/2$ and k_2 shows that in most cases, k_2 is larger than $k_1/2$, suggesting that the binding of the second monomer may be facilitated by the presence of a monomer already bound. Thus for Cre, protein- protein interactions between the bound and the incoming monomer may contribute to the cooperativity of binding. (Note however, that this reasoning is based on the assumption that for Cre, $k_{1a} = k_{1b}$, a point which cannot be determined on the basis of the present data.) The difference in the association rates of the first and second Cre monomers is not alone sufficient to account for the cooperativity of binding. Similarly to FLP, the Cre dimer complex has a slower dissociation rate than the monomer complex, showing that binding is stabilised by the presence of both monomers.

In summary, these results show that for both FLP and Cre, the cooperativitiy of binding is mainly mediated by a slow dimer dissociation rate. The fact that Cre has a higher DNA binding affinity than FLP is due both to a faster association rate, and to a slower dissociation rate for both first and the second monomer.

4.2.9 The kinetics of SM and SM2 formation are highly interdependent.

What are the relative contributions of the rate constants k_1, k_{-1}, k_2 and k_{-2} to the observed binding kinetics of the monomer and the dimer complex for FLP and Cre? To answer this question, time course curves were simulated in which each of the parameters was varied individually. Figure 4.13 shows the effect of variations in the parameters k_1, k_{-1}, k_2 and k_{-2} on the simulated binding kinetics of species SM and SM2 for FLP and Cre. Each parameter was increased two fold, five fold and tenfold in progressive steps.

Figure 4.13 shows that variation of a single parameter affects the binding kinetics of both species SM and SM2 in all cases. For example, for FLP, a two fold increase in k_1 (figure 4.13a, dotted line), causes an increase in the initial rate of formation of the species SM. At the same time, an increase in the rate of formation of the species SM2 is observed. This result can be explained by the fact that SM2 is formed from SM (equation 8), therefore an increase in the concentration of SM allows SM2 to be formed more quickly.

For Cre, a two fold increase in k_1 (figure 4.13e, dotted line) increases the concentration of the species SM, but has no effect on the species SM2. In contrast to FLP, a five fold and ten fold increase in k_1 has the effect of *reducing* the species SM2 (figure 4.13e, dashed and long dashed lines). This result can be explained by the fact that the concentration of Cre protein, 0.2 nM, is limiting in this experiment. Thus when the concentration of the species SM is increased, e.g., by a ten fold increase in k_1, to 0.13nM (figure 4.13e, long dashed line) then the concentration of free protein monomer is reduced to such an extent that the species SM2 is formed more slowly.

Increasing k_{-1}, the dissociation rate of the first monomer, results in a decreased concentration of both species SM and SM2 for FLP and Cre. (figures 4.13 b and f). This is due to the fact that the rate of SM2 formation is limited by the available concentration of SM.

Increasing k_2, the association rate of the second monomer, has the effect of increasing the concentration of SM2, at the expense of SM, whose concentration is reduced (figures 15c and g). The increase in SM2 is more pronounced for FLP than for Cre, again an effect of the limiting protein concentration in the Cre experiment.

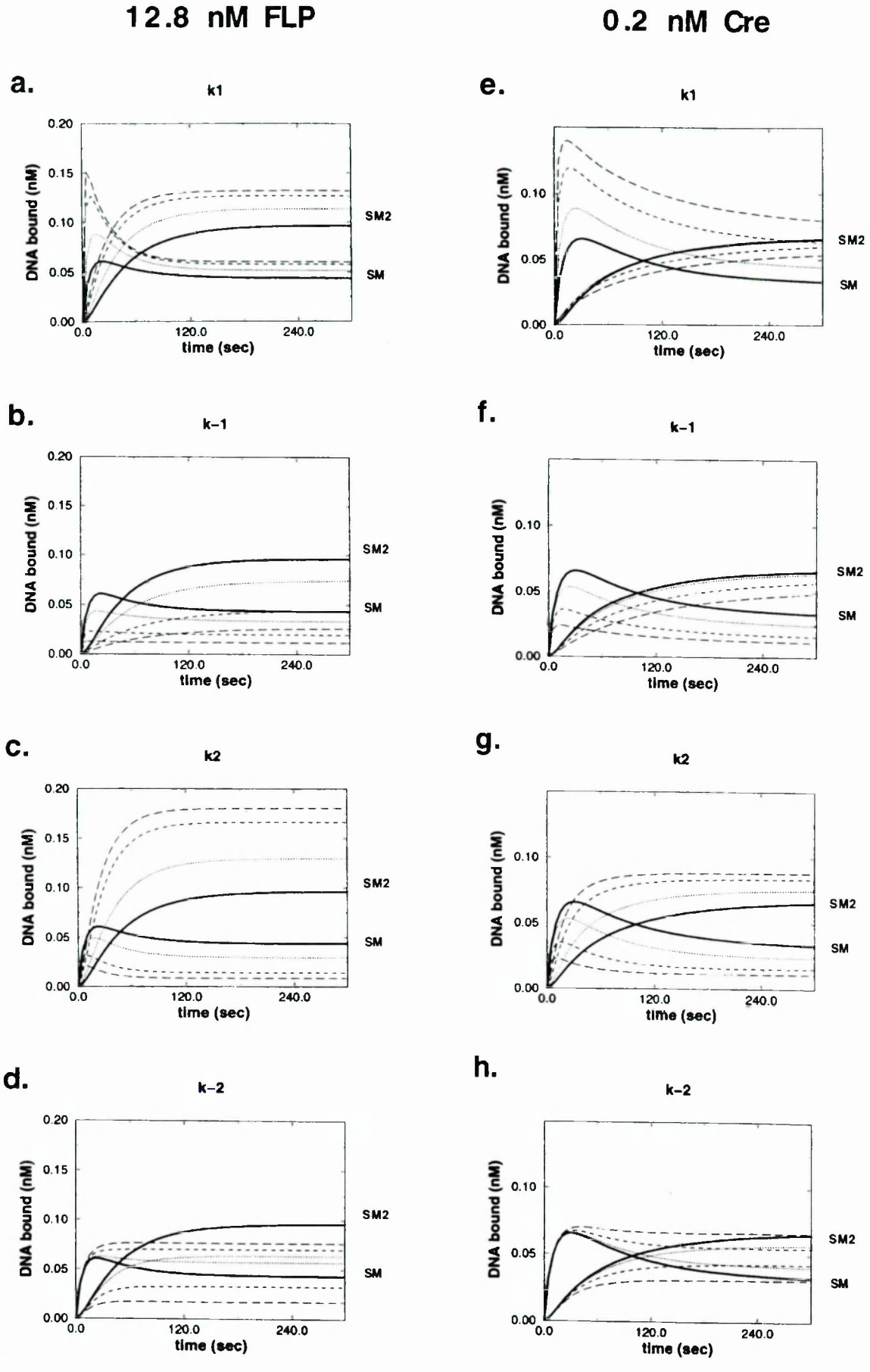
Increasing the value of k_{-2}, the dissociation rate of the second bound monomer, has the effect for both FLP and Cre, of decreasing the concentration of SM2, and thereby increasing the concentration of SM. The species SM2 dissociates to give SM plus free monomer, therefore an increase in SM2 dissociation gives rise to an increase in the concentration of the species SM. This effect is more pronounced for FLP than for Cre, (compare the dotted lines, showing two fold increase in k_{-2}, in figures 4.13 d and h). This result may be due to the relationship between k_2 and k_{-2} for each protein. For Cre, k_2/k_{-2} is greater than k_2/k_{-2} for FLP (Table 4.3). Thus an increase in the dissociation rate has less effect on the overall binding for Cre than for FLP.

In summary, the result of varying single binding parameters shows that the kinetics of monomer and dimer binding are highly interdependent, and illustrates the usefulness of the mathematical model as a tool for studying the exact nature of this interdependence in specific cases.

Figure 4.13 Effect of varying individual rate constants on simulated time course curves.

The optimised parameters for FLP at 12.8 nM and Cre at 0.2nM (figures 4.11(d) and (e), Table 4.3) were used to generate simulated time course curves for both proteins, shown as *heavy solid lines.* The curves are marked SM2 (complex 2) and SM (complex 1). Each of the four rate constants was then varied individually. In each case, the effect of increasing a single parameter in three progressive steps is shown as: two fold increase, *dotted line;* five fold increase, *dashed line*; ten fold increase, *long dashed line.* **(a)** FLP, effect of varying k_1. **(b)** FLP, effect of varying k_{-1} **(c)** FLP, effect of varying k_2. **(d)** FLP, effect of varying k_{-2}. **(e)** Cre, effect of varying k_1. **(f)** Cre, effect of varying k_{-1} **(g)** Cre, effect of varying k_2. **(h)** Cre, effect of varying k_{-2}.

Figure 4.13


12.8 nM FLP
0.2 nM Cre
a.
k1
b.
k-1
c.
k2
d.
k-2
e.
k1
f.
k-1
g.
k2
h.
k-2
DNA bound (nM)
time (sec)
SM2
SM
0.0
120.0
240.0
0.00
0.05
0.10
0.15
0.20

4.3.1. Comparison of FLP and Cre recombination *in vitro*: FLP recombines 100% of its substrate whilst Cre recombines 70%.

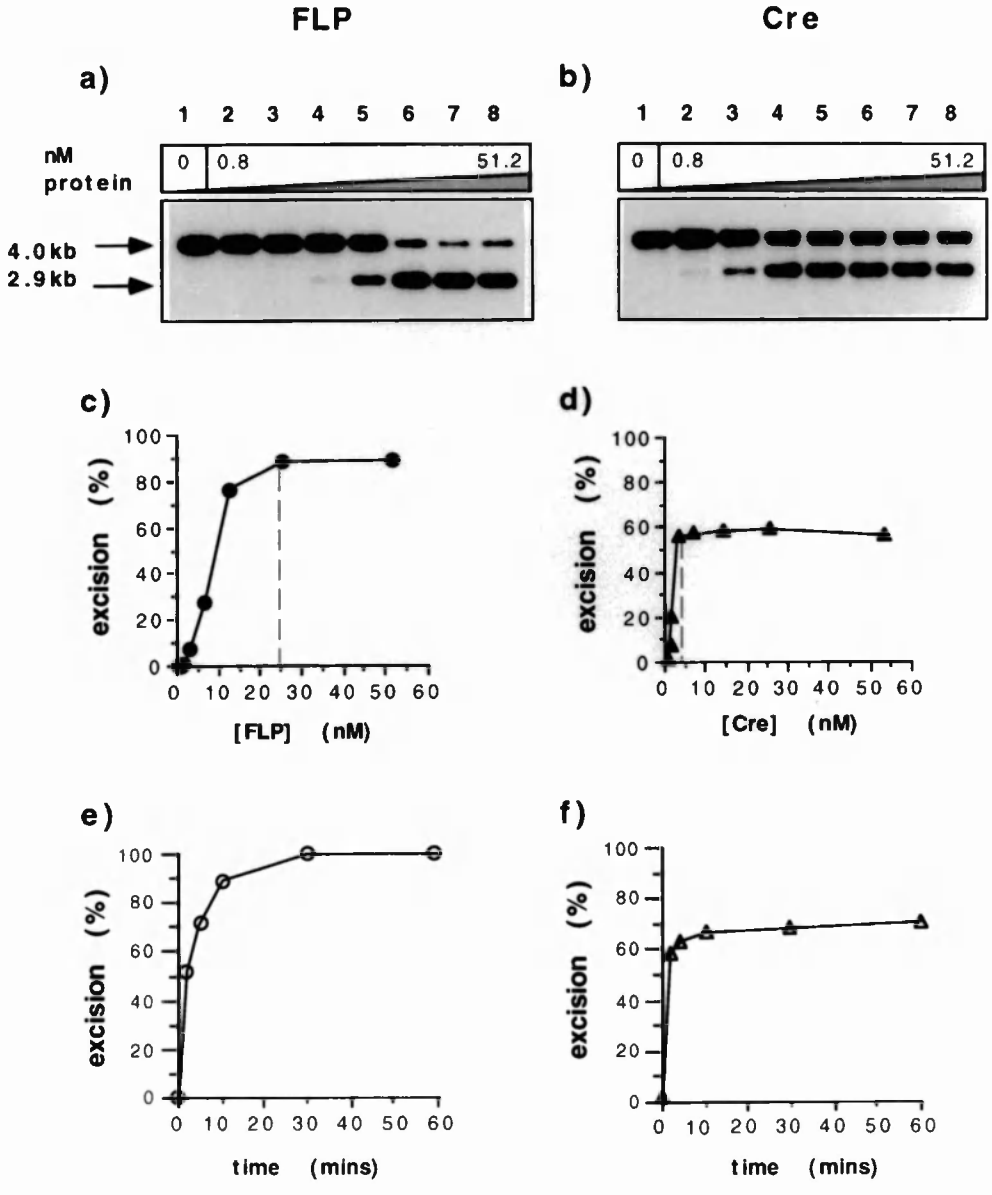
The last section defined in detail the equilibrium and rate constants for DNA binding using purified FLP and Cre proteins. Ideally, for a full kinetic description of the recombination reaction, rate constants for all of the steps shown in figure 4.1 would have to be measured. The steps of synapsis (k_3, k_{-3} in figure 4.1), the multiple steps of catalysis (simplified as k_4, k_{-4} in figure 4.1), and the dissociation of the synaptic complex (k_5, k_{-5} in figure 4.1) are not easily dissected. I reasoned that an examination of the kinetics of the complete reaction for FLP and Cre, in combination with accurate information for the initial steps, may throw some light on these intermediate steps.

In this section, the kinetics of FLP and Cre mediated recombination are compared using the linear excision substrates pSVpaZ (FLP) and pSVpaX (Cre). The excision assay is essentially as described for Kw in Chapter 3, and is outlined in figure 3.1. Conditions were determined in which both proteins recombine optimally. The effect of salt on purified FLP and Cre was examined. Both proteins tolerate a wide range of NaCl concentrations without loss of recombination activity (between 75 and 250mM tested, in the presence of 2mM $MgCl_2$), although an increase in intermolecular recombination was observed for both FLP and Cre at 250mM NaCl. (data not shown) This result is in agreement with that of Meyer Leon et al., (1987) using FLP purified in the same way as the present preparation. No evaluation of the effect of variations in monovalent cation concentration on Cre recombination activity has been published, although $MgCl_2$ has been shown to be necessary for optimal recombination (Abremski and Hoess, 1984). All subsequent experiments were carried out at 180mM NaCl, 2mM $MgCl_2$. Both recombinases were found to recombine at optimal levels in TAPS

Figure 4.14 Kinetics of recombination: purified FLP and Cre.

a, b) Protein titration. Varying amounts of pure FLP (a) or Cre (b) proteins were incubated with 0.4nM of the appropriate excision substrate for 60 minutes at 30°C (FLP) or 37°C (Cre). Unrecombined (4.0kb) and recombined (2.9kb) bands are shown on the left. Lane 1, no protein; lane 2, 0.8 nM; lane 3, 1.6 nM; lane 4, 3.2 nM; lane 5, 6.4 nM; lane 6, 12.8 nM; lane 7, 25.6 nM; lane 8, 51.2 nM. **c), d).** Quantification of a, b). % excision was plotted against protein concentration. On each plot, the *dashed line* indicates the minimum concentration of protein sufficient to give maximum recombination. **c),** *closed circles,* FLP. **d)**, *closed triangles,* Cre. **e), f). Time course of FLP and Cre recombination.** Protein concentrations were determined from c) and d) above as indicated by the dashed line. 25.6 nM FLP (e) or 3.2 nM Cre (f) were incubated with 0.4 nM of the appropriate excision substrate at the appropriate temperature, and aliquots removed at the time indicated on the plot. (e), *open circles*, FLP (% excision). (f) *open triangles*, Cre (% excision).

Figure 4.14


FLP
Cre
a)
b)
1 2 3 4 5 6 7 8
nM protein
0
0.8
51.2
4.0kb
2.9kb
c)
d)
excision (%)
100
80
60
40
20
0
0 10 20 30 40 50 60
[FLP] (nM)
[Cre] (nM)
e)
f)
time (mins)

buffer pH8.0. Other buffer components (PEG, glycerol, BSA, see "Materials and Methods") have been reported to enhance the stability of purified FLP (Gates and Cox, 1988), and were included in the reaction buffers for both proteins.

Under these conditions, the optimum ratio of protein monomer to target site was examined. For FLP at DNA concentrations of 0.2 nM and above, maximum recombination was observed at a protein monomer to half site ratio of between 10:1 and 16:1 (figure 4.14a). This result is in good agreement with previous determinations of the optimum ratio using FLP purified in the same way (Meyer-Leon et al., 1987, Gates and Cox, 1988). At DNA concentrations below 0.2 nM, a greater excess of FLP protein was necessary to achieve efficient recombination (not shown). For Cre under these conditions, the optimum protein monomer to half site ratio was approximately 2:1 (figure 4.14b). This is lower than the optimum ratio (6:1) reported by Abremski and Hoess (1984) for Cre purified by the same procedure. It is possible that the present preparation has a higher activity than that of Abremski and Hoess, or that the stabilising agents included in the buffer enhance Cre mediated recombination. In contrast to FLP, the efficiency of Cre mediated recombination was not adversely affected at DNA concentrations below 0.2 nM (data not shown), however, in order to compare the two recombinases systematically under the same conditions, the substrate concentration used in all experiments with purified protein was 0.2 nM or above. The effect of temperature on FLP and Cre has been characterised (Buchholz et al, (1996), see also Chapter 5), showing that Cre recombines optimally *in vitro* at 37°C, whereas the optimum temperature for FLP is 30°C. Reactions were therefore incubated at 30°C and 37°C for FLP and Cre respectively.

Comparison of the maximal levels of recombination reached by FLP and Cre (figure 4.14c, d) shows a striking difference between the two enzymes. FLP recombines 90% of the excision substrate in this experiment. (In other experiments, recombination reaches 100%, see for example, figure 4.14 e). For Cre on the other hand, the maximum level of excision reached is 60%. In other experiments, this level varies between 50 % and 75%, depending on the protein: substrate ratio, but more than 75% excision has not been observed. This result

is in agreement with the maximum level of Cre mediated excision reported by Abremski and Hoess (1984). The maximum level of FLP and Cre recombination is not affected by higher concentrations of protein (data not shown).

From the protein titration experiment, the optimum protein monomer: substrate ratio was defined for FLP and Cre (figure 4.14c, d) and a time course of recombination was carried out at this ratio (figure 4.14 e, f). FLP reached maximum excision of 100% in 30 minutes. For Cre, the reaction was almost complete within 2 minutes (60% excision), and reached the maximum of 67% in 10 minutes. The initial rate of recombination is not significantly different for the two proteins in this experiment (compare the 2 minute time point in figures 4.14e and f). Therefore the longer time taken for FLP to reach its maximum reflects the higher level of total recombination achieved. In summary, the comparison of FLP and Cre shows that FLP can recombine 100% of an excision substrate, whereas Cre recombines 70%, and that both recombinases have a similar initial rate of excision.

4.3.2 Estimation of the fraction of active protein in FLP and Cre preparations.

The experiments presented in figure 4.14 raise two questions. Firstly, why is more FLP than Cre required to achieve maximum recombination, and secondly, why is the maximum level reached lower for Cre than for FLP? This section is concerned with the first question; the second question is addressed in section 4.3.6

Considering the first question, there are at least two possible explanations. Firstly, the lower DNA affinity of FLP for the FRT site may explain the requirement for more protein. The excision substrate has to be bound by four protein monomers in order to undergo recombination (figure 4.1). The lower the affinity of the protein for its substrate, the higher the concentration required to occupy all four binding sites simultaneously for long enough for synapsis to take place.

A second possible explanation is that the FLP preparation is not fully active, so that the calculation of active monomers in the FLP preparation is an overestimate. This point is crucial not only to the analysis which follows, but also to the entire preceding section, in which the rate constants for DNA binding are calculated on the basis of the concentration of active monomer. For Cre, the issue of protein activity can be addressed by inspection of the gel mobility shift titration experiments described in section 4.2.1, figure 4.4. Under conditions in which the concentration of Cre protein is low relative to the concentration of substrate, (e.g., figure 4.4 lane 3), calculation of the total monomer bound shows that 100 % of the monomer is bound to DNA as dimer or monomer complexes. This result suggests that the Cre protein preparation is close to 100% active. The same calculation for FLP (figure 4.3) shows that the highest proportion of FLP monomer bound at any one time is 10%. Again the dilemma arises of whether this result reflects a true low affinity, or a low activity. Discussing the possibility that inactive protein may contribute to the requirement for a large molar excess of FLP over 1/2 site for efficient recombination, Gates and Cox (1988) pointed out that the FLP preparation used was purified by affinity chromatography on a site specific DNA Sepharose column (Meyer-Leon et al, 1987). The FLP preparation used here was purified in the same way. This procedure may be expected to select for FLP monomers active in DNA binding.

The possibility remains, however that some of the FLP preparation may be inactive. To address this problem further, I studied the effect of substrate concentration on recombination kinetics of FLP and Cre. At a given protein concentration, the concentration of excision substrate was varied and recombination was observed over a time course. Central to this approach is the prediction that above a certain substrate concentration, recombination should be inhibited. This should happen because of the co-operative nature of the reaction: all four half sites of the substrate have to be occupied for recombination to take place (figure 4.1). In the extreme case, if the number of half sites exceeds the number of active monomers available, then the protein will be "diluted" amongst the substrate molecules, each substrate being bound by, on average, less than one monomer. Recombination would be expected to be reduced under

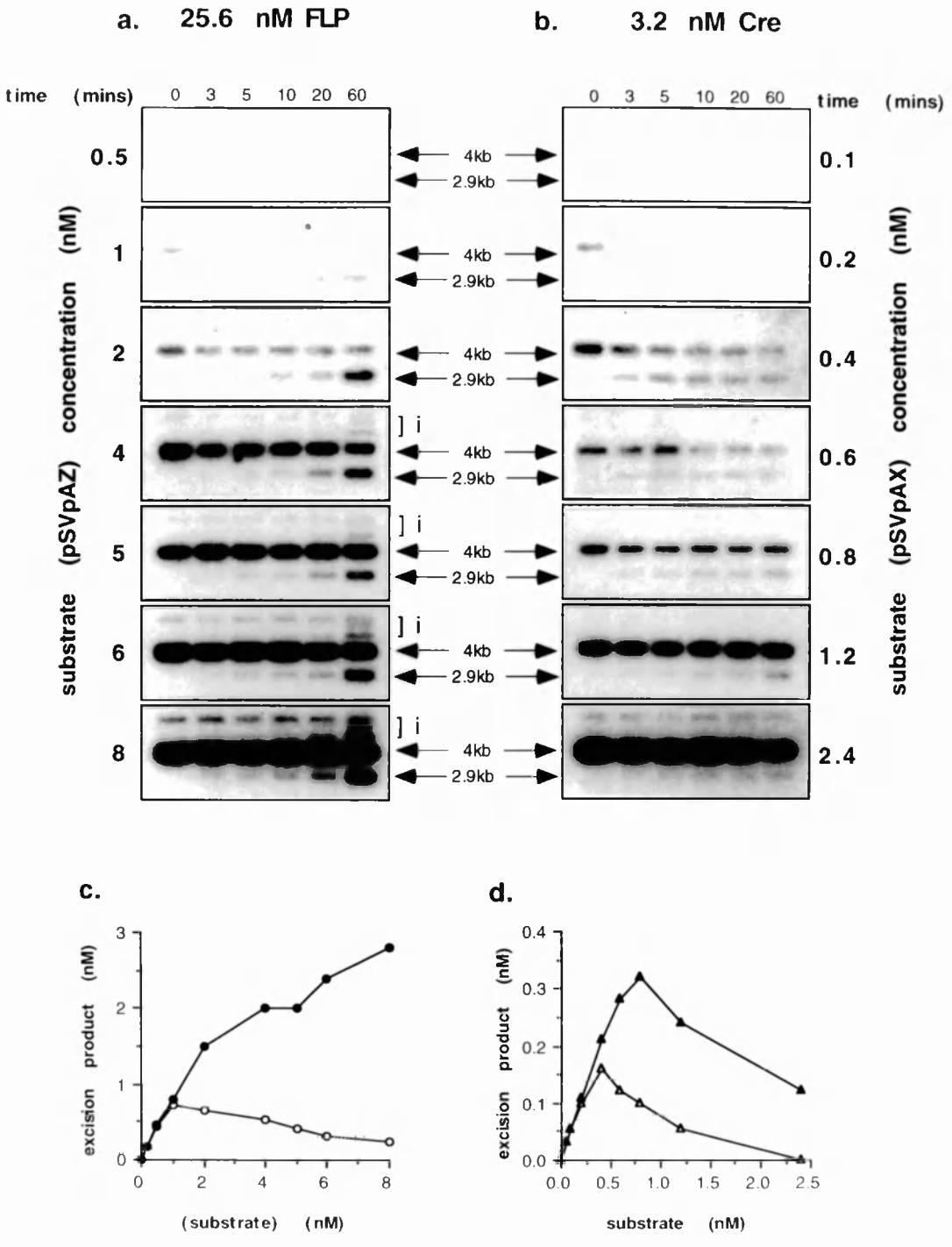
these conditions. I reasoned therefore, that the substrate concentration at which recombination is inhibited may give an indication of the true concentration of active monomer.

Figure 4.15 shows the substrate saturation behaviour of FLP and Cre. The protein input was determined from figure 4.14 as the minimum amount required for maximum recombination of 0.4nM substrate (25.6nM FLP, and 3.2 nM Cre). Strikingly, the total amount of Cre recombination observed at 60 minutes reaches a maximum of 0.32nM at 0.8 nM substrate input, and is severely reduced at higher substrate concentrations, (figure 4.15d, closed triangles). In contrast, for FLP, the total amount of recombination increases approximately linearly over the substrate concentrations tested (up to 8nM, figure 4.15c, closed circles). The maximum excision product observed is 2.8 nM. The initial rate of Cre mediated recombination increases linearly with substrate concentration and reaches a maximum at 0.4 nM substrate, of 0.16 nM excision product in 2 minutes (figure 4.15d, open triangles). The initial rate is strongly reduced by each further increase in substrate concentration. For FLP, the initial rate also increases linearly with substrate concentration and reaches a maximum at 1 nM substrate, of 0.7 nM excision product in 2 minutes. For FLP, the initial rate is slightly reduced by each further increase in substrate concentration (figure 4.15c, open circles). The protein monomer to half site ratio at which maximum initial recombination is reached is 2:1 for Cre, and 6.4 : 1 for FLP. The protein monomer to half site ratio at which maximum total recombination is reached is 1:1 for Cre. For FLP, at 0.8: 1 monomer to half site ratio, the total recombination observed in 60 minutes is still increasing.

Comparison of the effect of substrate concentration on the initial rate of recombination for FLP and Cre suggests that there is indeed more active FLP than Cre in this experiment. To calculate exactly how much FLP is present, the following reasoning can be used: if the DNA affinities and all other characteristics of the two proteins were identical, then the maximum initial rate of recombination observed would give an estimate of the activity of the preparation. For 25.6 nM FLP, the maximum initial rate of recombination observed is 0.75 nM excision product in 2 minutes (figure 4.15c,

Figure 4.15 Effect of varying the substrate input: purified FLP and Cre. a) 25.6 nM FLP was incubated at 30°C with the following concentrations of excision substrate: 0.2 nM (not shown) 0.5 nM, 1 nM, 2 nM, 4 nM, 5 nM, 6 nM, and 8 nM, as indicated at the left of the figure. Aliquots were removed from each reaction at the time points shown on the figure. **b)** 3.2 nM Cre was incubated at 37°C with the following concentrations of excision substrate: 0.05 nM (not shown) 0.1 nM, 0.2 nM, 0.4 nM, 0.6 nM, 0.8 nM ,1.2 nM, and 2.4 nM, as indicated at the right of the figure. Aliquots were removed from each reaction at the time points shown on the figure. Unrecombined (4.0kb) and recombined (2.9kb) bands are shown in the centre of the figure. Intermolecular recombination products are marked "i". The substrate used was unlabelled psVpaZ (FLP) or pSVpaX (Cre). DNA was visualised by southern hybridisation and quantified by PhosphorImager analysis. **c), d)**. Comparison of recombination after 3 minutes and 60 minutes in a) and b). Total nM excision product at each substrate input is plotted against substrate concentration. **c)** FLP. *Open circles*, nM excision product at 3 minutes. *Closed circles*, nM excision product at 60 minutes. **d),** Cre. *Open triangles,* nM excision product at 3 minutes. *Closed triangles,* nM excision product at 60 minutes.

Figure 4.15


a. 25.6 nM FLP
b. 3.2 nM Cre
time (mins)
0 3 5 10 20 60
substrate (pSVpAZ) concentration (nM)
substrate (pSVpAX) concentration (nM)
0.5
1
2
4
5
6
8
0.1
0.2
0.4
0.6
0.8
1.2
2.4
4kb
2.9kb
] i
c.
d.
excision product (nM)
(substrate) (nM)
substrate (nM)

open circles). For 3.2 nM Cre the maximum initial rate of recombination observed is 0.16 nM excision product in 2 minutes (figure 4.15d, open triangles). This corresponds to an initial excision rate of 0.014/minute for FLP, and 0.025/minute for Cre. If the characteristics of the two proteins were identical, and given that the Cre preparation is estimated to be 100% active, then this result indicates that 56% of the FLP preparation is active. This gives a lower limit to the estimation of FLP activity, since a lower activity would not be sufficient to explain the initial rates of recombination observed, unless the affinity of FLP for the FRT were higher than that of Cre for LoxP. However, in the previous section I showed that FLP has a lower specific DNA binding affinity than Cre. Even if only 56 % of the FLP preparation were active, the calculation of the rate constants k_1 and k_2 would still result in an approximately 50 fold higher specific DNA affinity for Cre than for FLP (see Table 4.3). I conclude from this reasoning that 56% is an underestimate of the active protein in the FLP preparation. Thus the maximum initial rate of FLP recombination observed, (0.014/minute, compared to 0.025/minute for Cre), may simply reflect the lower DNA binding affinity of FLP for its target. It is possible that the FLP preparation is 100 % active.

Other qualitative differences in the behaviour of FLP and Cre in the substrate saturation assay indicate that the two proteins have very different recombination kinetics. Most strikingly, for FLP, the total amount of recombination observed after 60 minutes increases with increasing substrate concentration, whilst the initial rate decreases. For Cre on the other hand, the initial and final levels of recombination are both sharply reduced with increasing substrate concentration. (Figure 4.15c and d). This behaviour may be due to the differences in DNA binding affinities of FLP and Cre for their substrates. It is also possible however, that this behaviour is caused by a difference in the non -specific DNA binding affinities of FLP and Cre. For instance, if Cre has a higher non specific affinity for DNA than FLP, then by increasing the total amount of DNA in a reaction, the amount of free Cre monomer would be effectively reduced, and recombination would be inhibited. This possibility was addressed by repeating the experiment with the inclusion of carrier DNA such that the total DNA concentration was equal in all

experiments, with similar results to those shown in figure 4.15 (not shown).

4.3.3 *In vitro* translated protein behaves similarly to purified protein.

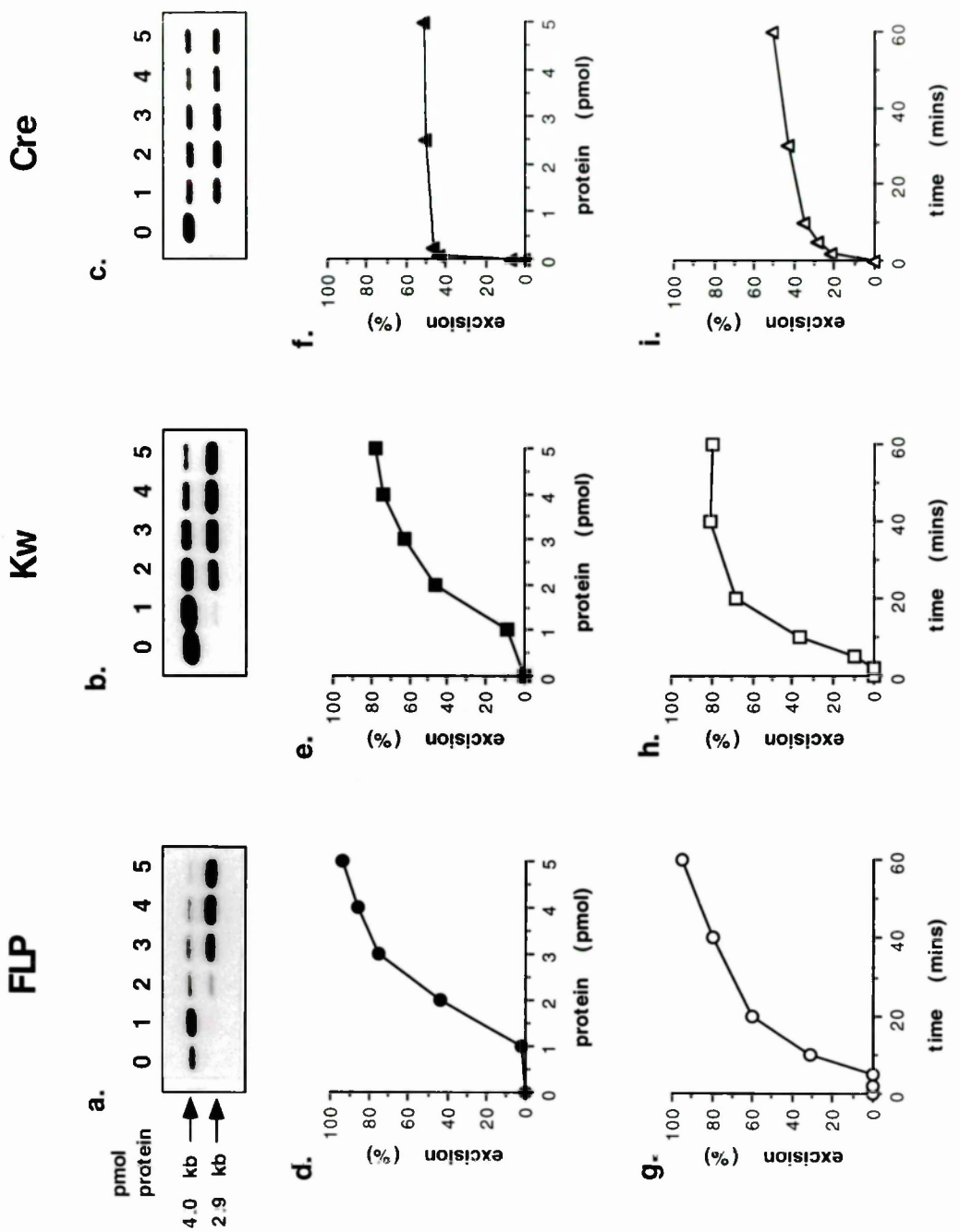
The analysis of recombination kinetics using purified protein enables accurate determination of kinetic parameters, because the number of active protein monomers can be relatively accurately determined. However, purified preparations are not always available. I was interested to see whether the characteristics of recombination displayed by purified FLP and Cre would also be reproduced by *In vitro* translated proteins, for two reasons. Firstly, so that the Kw protein (Chapter 3) which has not been purified, could be compared to FLP and Cre. Secondly, it would eventually be interesting to compare the properties of FLP and Cre mutants, whose production in small quantities by *in vitro* translation would be more convenient than purification.

Conditions were determined in which *in vitro* translated FLP, Cre and Kw proteins performed optimally. One significant difference between the *in vitro* translated and purified FLP preparations was that efficiency of recombination by *in vitro* translated FLP was not adversely affected by substrate concentrations below 0.2nM. (see section 4.3.1). I expect that this effect was due to factors in the rabbit reticulocyte lysate which stabilise FLP activity. *In vitro* translated FLP also recombined efficiently in the absence of BSA, PEG and glycerol. These components were omitted from the buffers of the experiments in this section.

In figure 4.16, recombination is compared for *in vitro* translated FLP, Kw and Cre proteins. The maximum levels of recombination achieved by *in vitro* translated FLP (a,d) and Cre (c, f) are similar to those observed for the purified proteins. (95% and 54% for FLP and Cre respectively). The maximum level reached for Kw (b, e) is 80%. (In other experiments, where the protein: substrate ratio was

Figure 4.16 Kinetics of recombination: *in vitro* translated FLP, Kw and Cre. a), b), c), Protein titration 1-5 pmol of *in vitro* translated FLP (a), Kw (b), or Cre (c) proteins were incubated in 20µl volume with 0.01 nM of the appropriate end labelled excision substrate for 60 minutes at 30°C (FLP, Kw) or 37°C (Cre). The volume of protein added was made up to 5µl in each case with mock translation mix as described in "Materials and Methods". Unrecombined (4.0kb) and recombined (2.9kb) bands are shown on the left. **d), e), f).** Quantification of protein titration data. **d)**, *closed circles*, FLP; **e)** *closed squares*, Kw, **f)** *closed triangles,* Cre. To make the Cre graph, additional experiments were done using 0.125, 0.25 and 0.5 pmol protein (data not shown). **g), h), i).** Time course of recombination. FLP, Kw or Cre protein was incubated with 0.01 nM of the appropriate end labelled excision substrate and aliquots removed at the times indicated on the figure. Protein concentrations were determined from d), e), and f) above as the minimum amount required to give maximum recombination in 60 mins., and reactions were scaled up to 80µl. **g)** *open circles*, FLP (20pmol). **(h)**, *open squares,* Kw (20pmol). **(i)**, *open triangles*, Cre, (2pmol).

Figure 4.16


FLP
Kw
Cre
a.
b.
c.
pmol protein
0 1 2 3 4 5
4.0 kb
2.9 kb
d.
e.
f.
g.
h.
i.
excision (%)
protein (pmol)
time (mins)

higher, the total Kw mediated recombination did not exceed 80%.) *In vitro* translated protein was quantified by 35S methionine incorporation as described in "Materials and Methods". The minimum amount of protein required to reach maximum recombination was determined for each protein. Figure 4.16 d, e, and f shows that for all three proteins, optimum recombination requires picomolar quantities of protein. This corresponds for all proteins to a molar excess of protein over half site of over 100. This result suggests either that the estimation of protein concentration by 35S methionine incorporation is inaccurate, or that the *in vitro* translated protein preparations contain a large proportion of inactive monomer.

For each protein, a time course of recombination was performed at the optimum protein concentration (figure 4.16 g,h,i). For FLP and Cre, the maximum recombination observed is similar to that observed with purified protein. For Cre (i), the initial rate is similar to that observed with purified protein, but for FLP (g), the kinetics of the reaction are slower. A 5 minute lag is observed in which no recombination is seen, and the maximum recombination observed, 92%, is reached only after 60 minutes. Kw (h) shows slightly faster recombination than FLP, with a two minute lag, reaching its maximum of 80% recombination in 40 minutes. I conclude that the kinetics of recombination by *in vitro* translated proteins are qualitatively similar, and that comparisons can be made between different proteins produced by *in vitro* translation.

4.3.4. Substrate saturation behaviour of *in vitro* translated proteins enables comparison of Kw with FLP and Cre.

To compare the behaviour of Kw to that of FLP and Cre in the substrate saturation assay, the experiment in figure 4.17 was carried out. Different concentrations of excision substrate were incubated with a given concentration of *in vitro* translated FLP, Kw or Cre protein, determined from figure 4.16 as the minimum concentration required to give maximum recombination of 0.01nM substrate in 60 minutes.

In vitro translated FLP and Cre behave very similarly to the purified proteins in this assay. Over the range of substrate concentrations tested, (0.000625 to 0.24 nM) the total FLP recombination observed at 120 minutes increases linearly with increasing substrate concentration (figure 4.17d). For Cre, the level of total recombination reaches a maximum at 0.015 nM substrate input, and decreases abruptly with increasing substrate concentration (figure 4.17 f). Kw behaves similarly to FLP, in that total recombination increases linearly over the range of substrate concentrations tested (figure 4.17e). Further similarities between the *in vitro* translated and pure proteins are evident upon comparison of the initial rates with the total recombination: For FLP, the total recombination increases linearly with increasing substrate concentration, whilst the initial rate begins to reach a plateau at the higher concentrations (figure 4.17g). For *in vitro* translated Cre, as for the purified protein, the initial and final levels of recombination are closely linked to substrate concentration: in this experiment, maximum total recombination and maximum initial recombination are both measured at the same substrate concentration (0.015 nM, figure 4.17 i). In terms of initial recombination rates, Kw behaves very similarly to FLP. The maximum initial recombination rate is somewhat faster than for FLP (0.23 and 0.16nM excision product/min for Kw and FLP respectively, figure 4.17g,h). The maximum initial rate of recombination for Kw is reached at a slightly lower substrate concentration than for FLP (0.2nM and over 2.4 nM for Kw and FLP respectively, figure 4.17 g,h). If the behaviour of FLP and Cre proteins in the substrate saturation assay is a reflection of their DNA binding affinities as discussed in section 4.3.2 above, then the behaviour of Kw in this assay suggests that its DNA binding affinity may be similar to, but slightly higher than that of FLP.

The curves in figure 4.17 g, h, and i, showing initial rate of recombination as a function of substrate concentration differ from those measured using purified protein in that the initial part of the curve has a sigmoid shape. The sigmoid shape has not been observed in experiments with purified FLP and Cre, even at substrate

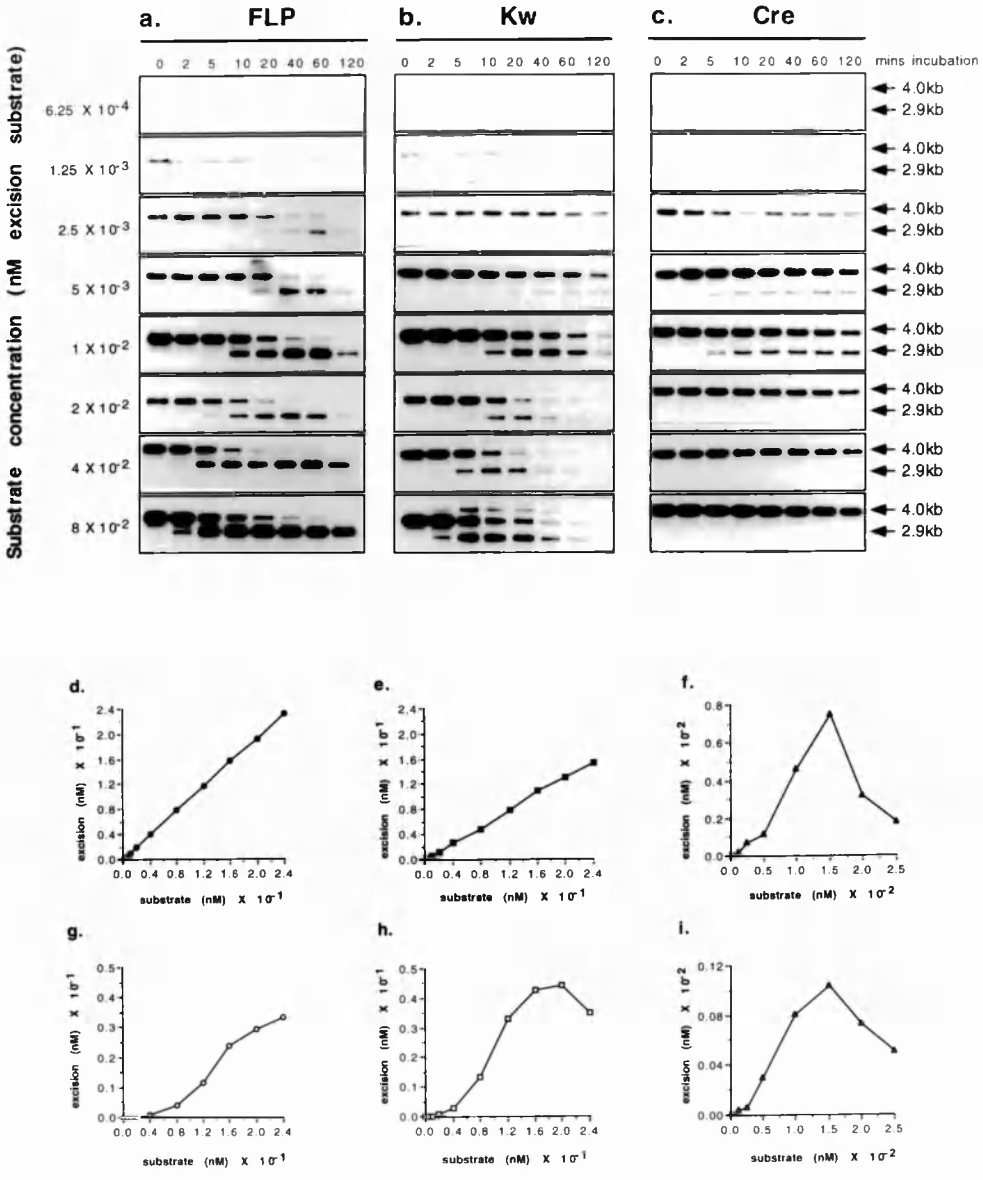
Figure 4.17 Effect of varying the substrate input: *in vitro* translated FLP, Kw and Cre.

The amount of protein used was determined from figure 4.16 as the minimum amount of protein required for maximum recombination of 0.01 nM substrate in 60 minutes, and reactions were scaled up to 80µl. **a)** 20 pmol FLP, **b)** 20 pmol Kw, **c)** 2 pmol Cre were incubated with different concentrations of the appropriate end labelled excision substrate as shown on the left of the figure. Aliquots were taken at the time intervals shown at the top of the figure.

d), e), f). Total recombination. The total nM excision product at 120 minutes was calculated from each reaction in a), b), c) and similar experiments (see below), and plotted against the substrate input concentration. To make the FLP and Kw graphs, extra experiments were carried out at substrate concentrations of 0.12 nM, 0.16 nM, 0.2 nM and 0.24 nM. To make the Cre graph, extra experiments were carried out at 0.015 nM, and 0.025 nM. **d)**, *closed circles*, FLP. **e)**, *closed squares*, Kw. **f)** *closed triangles*, Cre.

g), h), i). Initial recombination. The total nM excision product at 2 minutes was calculated from each reaction in a), b), c) and the additional experiments described for d) e) and f), and plotted against the substrate input concentration. **g)**, *open circles*, FLP. **h)**, *open squares*, Kw. **i)** *open triangles*, Cre.

Figure 4.17


a. FLP
b. Kw
c. Cre
0 2 5 10 20 40 60 120
mins incubation
Substrate concentration (nM excision substrate)
6.25 X 10-4
1.25 X 10-3
2.5 X 10-3
5 X 10-3
1 X 10-2
2 X 10-2
4 X 10-2
8 X 10-2
4.0kb
2.9kb
d.
e.
f.
g.
h.
i.
excision (nM) X 10-1
excision (nM) X 10-2
substrate (nM) X 10-1
substrate (nM) X 10-2

concentrations similar to those used in figure 4.17. The sigmoid shape observed with *in vitro* translated protein is unexpected. In situations where the protein is in excess over the substrate, a linear increase of initial recombination with increase in substrate input would be expected. (Note that the sigmoid shape observed in similar experiments with co-operative systems such as haemoglobin binding to oxygen depends on the fact that the substrate molecules are independent of each other (Ebenhöh, 1975)). Therefore a process which is related non-linearly to substrate concentration must be responsible for the sigmoid curves observed. One such process could be for example, the binding of more than two monomers at each inverted repeat, thereby blocking the formation of a productive synapse. Such a process may occur preferentially in *in vitro* translated preparations. Alternatively, inactive monomers, or indeed components of the lysate itself, may participate in such an aggregation. Since this behaviour is not observed with purified proteins, it is not discussed further here.

In summary, the substrate saturation assay shows that *in vitro* translated FLP and Cre behave similarly to the purified proteins in most respects, and shows that Kw behaves similarly to FLP, suggesting that the DNA binding affinities of FLP and Kw are probably similar.

4.3.5 Further characterisation of purified FLP and Cre recombination: plasmid cointegration.

The experiments described so far are all based on an assay for excision, in which the two inverted repeats are separated by 1.1kb. This distance may be expected to have an effect on the ease with which two bound inverted repeats on the same molecule can find each other, since it effectively constrains the volume in which they can diffuse. If the two target sites are on separate molecules, then this constraint is removed, and the volume in which two bound sites must diffuse in order to find each other is increased to that of the reaction itself. Comparison of the kinetics of such an intermolecular reaction with those of the excision reaction studied

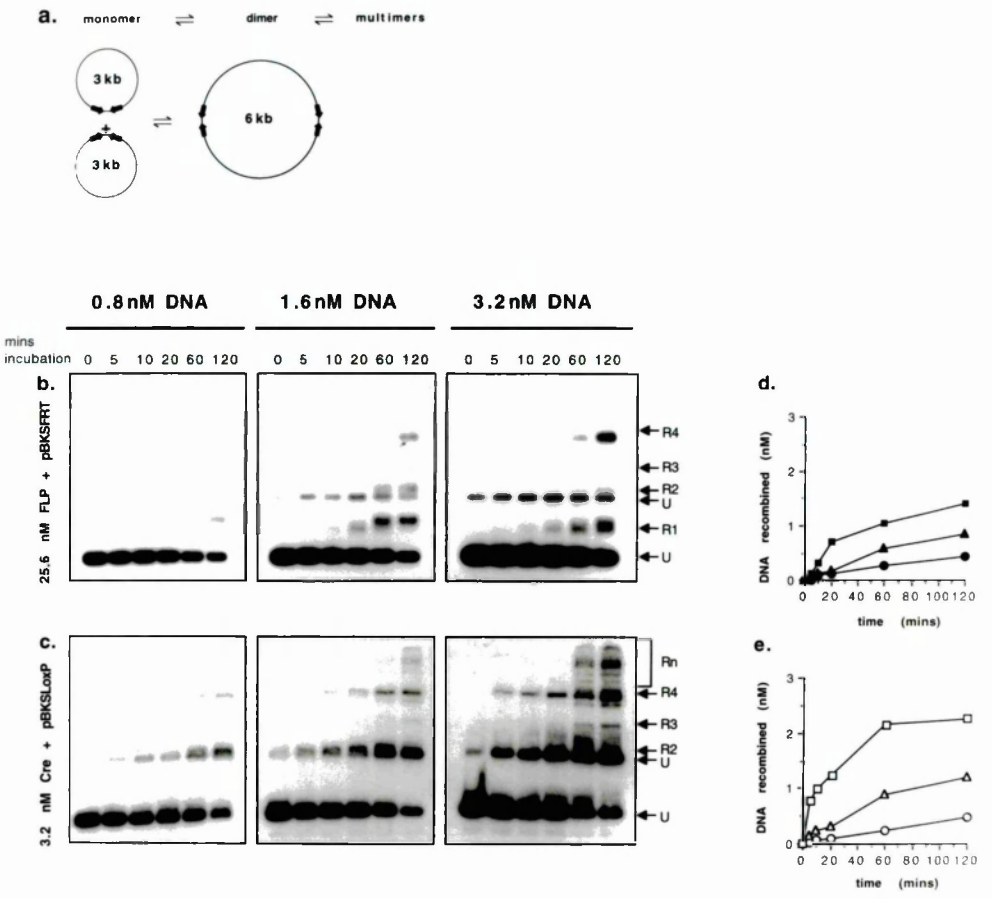
so far may give some idea of the parameters governing synapsis for FLP and Cre.

To investigate this idea, I designed the plasmid cointegration assay shown in figure 4.18a. The assay is based on a 3kb circular plasmid bearing a single inverted repeat FRT (pBKSFRT) or LoxP site (pBKSLoxP). Recombination between two plasmids results in a dimer. Further intermolecular recombination events will give plasmid multimers. Because recombination is reversible, excision can also result in resolution to smaller molecules. Supercoiled pBKSFRT or pBKSLoxP was incubated with FLP or Cre at three different substrate concentrations. The lowest concentration, 0.8nM, was chosen so that the protein monomer to half site ratio was equivalent to that at which optimum excision is observed (16:1 for FLP, 2:1 for Cre). At this DNA concentration, both FLP and Cre multimerise the plasmid at a similar rate, which is slower than that observed for the excision assay. About 0.5nM of the plasmid is multimerised after 2 hours in both cases (figure 4.18 b,c, 0.8nM DNA). This quantification is very probably an underestimation of the total number of intermolecular events, since at equilibrium, plasmid multimers are likely to be excised faster than they are formed. Initial reaction rates, where integration is prevalent over excision, may give a better indication of the true rate of synapsis for FLP and Cre. At 0.8nM DNA the initial rate is too slow to measure, but increases with increasing substrate concentration for both FLP and Cre. At the highest substrate concentration, 3.2 nM, Cre produces 0.75nM multimer in the first five minutes, (figure 4.18e) whereas FLP produces 0.2 nM multimer in 5 minutes (figure 4.18d). Note that in intramolecular recombination assays, substantial excision occurs within the first five minutes (see for example, figure 4.14e, f), suggesting that the multimers observed in 5 minutes in figure 4.18 do not solely result from intermolecular events. However, since the initial rates of FLP and Cre mediated excision are similar (see figure 4.14 and section 4.3.1) this result shows that under these conditions, Cre carries out intermolecular recombination faster than FLP. This suggests that the rate of synapsis may be faster for Cre than for FLP.

Figure 4.18 FLP and Cre intermolecular recombination

a) Recombination substrates and products. Recombination target sites are represented as *black arrows*. Multimerisation of the 3kb plasmid produces larger molecules. See text for full description of assay. **b), c).** Southern blot of intermolecular recombination assay. RNA probes used for Southern blotting correspond to 800 base pairs covering the FRT or LoxP site of the substrate plasmid. **b)** 25.6nM FLP was incubated with different amounts of pBKSFRT as shown above the figure. Aliquots were removed at different times as indicated. Unrecombined bands are labelled "U" on the right of the figure. Recombined products are labelled "R1" to "R4". **c).** 3.2nM Cre was incubated with different amounts of pBKSLoxP and aliquots removed at the times shown. Unrecombined bands are labelled "U" and recombined products are labelled "R2" to Rn". FLP and Cre recombination products which comigrate are given the same name. **d), e).** Quantification of **b)** and **c).** nM DNA recombined is plotted against time for each of the input DNA concentrations as follows: **d)** FLP, *closed circles*, 0.8 nM DNA; *closed triangles,* 1.6 nM DNA; *closed squares,* 3.2 nM DNA. **e)** Cre, *open circles*, 0.8 nM DNA; *open triangles,* 1.6 nM DNA; *open squares,* 3.2 nM DNA. nM DNA recombined was calculated for each lane using PhosphorImager analysis as (fraction of total counts in recombined bands) X (total nM input DNA in experiment).

Figure 4.18


a.
monomer
dimer
multimers
3 kb
3 kb
6 kb
0.8nM DNA
1.6nM DNA
3.2nM DNA
mins incubation
0 5 10 20 60 120
b.
25.6 nM FLP + pBKSFRT
R4
R3
R2
U
R1
U
c.
3.2 nM Cre + pBKSLoxP
Rn
R4
R3
R2
U
U
d.
DNA recombined (nM)
time (mins)
e.
DNA recombined (nM)
time (mins)

The stimulatory effect of substrate concentration on both FLP and Cre in this assay is in contrast to the effect of substrate concentration on the two recombinases in the excision based substrate titration assay (figure 4.15). This can be explained by the fact that in the cointegration assay, the two target sites are not linked on the same molecule, meaning that the effect of "protein dilution" is greatly reduced. In addition, the rate of synapsis between separate molecules would be increased by increasing substrate concentration. The fact that increasing substrate concentration has a greater effect on Cre than on FLP, is consistent with the idea that Cre may have a faster rate of synapsis than FLP.

Interestingly, in the only experiment in which recombination reached an equilibrium (Cre, 3.2nM substrate) the maximum recombination is 65%, which corresponds to the maximum recombination observed in excision assays with Cre. Further experiments with a range of substrate: protein ratios would be necessary to determine whether this is significant.

Additionally, some interesting qualitative differences between FLP and Cre emerge from the plasmid cointegration experiment. Firstly, although certain multimers formed by the two proteins are identifiable as probably the same species (R2, R3 and R4 in figure 4.18 b and c), other species are formed only by one or other protein. For example the band marked "R1" in figure 4.18b, is only formed by FLP. This may be a catenated form, caused by excision on a supercoiled plasmid multimer. It has been shown that FLP gives catenated products upon recombination of supercoiled excision substrates (Beatty et al., 1986), whereas Cre gives mostly unlinked circles (Abremski and Hoess, 1985). In addition, the many slowly migrating products marked "Rn" in figure 4.18b are only observed with Cre and not with FLP. These may represent different supercoiled forms (Cre has been shown to have a Topoisomerase I like activity, (Abremski et al., 1986(a) and (b)). In addition, some of these products may represent higher multimers of the plasmid pBKSLoxP.

In summary, the plasmid cointegration assay shows that for both FLP and Cre, the initial rate of intermolecular recombination is

slower than the initial rate of excision for equivalent protein: substrate ratios. At higher substrate concentrations, Cre carries out integration faster than FLP, suggesting that Cre may have a faster rate of synapsis than FLP.

4.3.6 Mathematical modelling of FLP and Cre recombination: Determination of unknown parameters

Three recombination assays have been used in this chapter: the excision time course, the substrate titration, and the intermolecular recombination time course. In all three assays, FLP and Cre demonstrate significant quantitative and qualitative differences. What is the reason for these differences? Are they entirely due to the differences in DNA binding affinity measured in the first part of this chapter, or do other parameters play a role?

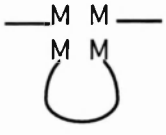
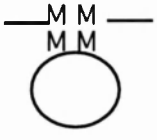
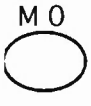
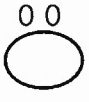
To answer these questions, I decided to take a mathematical approach. A mathematical model was constructed on the basis of classical kinetic equations as described in "Materials and Methods" (section 2.10.4), describing each of the species shown in figure 4.1, and Table 4.4. The rate of change of each species with respect to time is described by differential equations in terms of the species from which it is made, the species into which it can be converted, and the rate constants governing these processes. The model is incorporated into a program which solves the differential equations and gives simulated data for the evolution of each species over time. The model describes a general excision recombination reaction in which four protein monomers are required to reversibly recombine a single substrate, giving two products. The values of all rate constants, and of protein and substrate input can be varied to represent specific cases.

The rate constants k_1, k_{-1}, k_2 and k_{-2} as measured for FLP and Cre in sections 4.2.1 to 4.2.9 can be put directly into the model. From figure 4.1 and Table 4.4 it can be seen that this leaves three pairs of unknown rate constants: k_3 and k_{-3}, the back and forward rate of synapsis respectively; k_4 and k_{-4}, the back and forward rate of

Table 4.4. Complexes described in Mathematical model for recombination.

The model starts with free protein monomer and free substrate. The substrate has two full recombinase binding sites, each of which has two half sites, (designated 1 to 4 in the scheme below). Each half site can be bound by one protein monomer. The intrinsic affinities of monomer for half sites 1, 2, 3, and 4 are assumed to be identical. A half site is designated 0 if unoccupied, and M if occupied.

Table 4.4 Complexes described in mathematical model for recombination

Name	Configuration: 1/2 site __1 2_____3 4__	Description	Rate constants
M		free monomer	)
S	__0 0_____0 0__	free substrate	)
SM	__M 0_____0 0__	one monomer bound	)
SM2A	__M 0_____M 0__	one monomer at each site	)k_1,k_{-1}
SM2B	__M M_____0 0__	two monomers at one site	)k_2,k_{-2}
SM3	__M M_____M 0__	three monomers bound	)
SM4	__M M_____M M__	four monomers bound	)
I		synaptic complex	k_3, k_{-3}
IEP		excised synaptic complex	k_4, k_{-4}
EPM2	__M M__	linear excision product with two monomers bound	k_5, k_{-5}
LPM2		circular excision product with two monomers bound	k_5, k_{-5}
EPM	__M 0__	linear excision product with one monomer bound	k_{-2}, k_2
LPM		circular excision product with one monomer bound	k_{-2}, k_2
EP	__0 0__	free linear product	k_{-1}, k_1
LP		free circle	k_{-1}, k_1

recombination within the synapse; and k_5 and k_{-5}, the dissociation and reassociation rate of the synaptic complex after recombination. The question to be asked by the mathematical model is thus, whether there is a unique set of these parameters which, in combination with the measured values of the DNA binding parameters k_1, k_{-1}, k_2 and k_{-2} for FLP and Cre, can describe the observed time course and substrate titration behaviour of the two proteins. If such a set of parameters can be found, then how do they differ for the two proteins?

Before embarking on the search for the right combination of parameters, I decided to further simplify the task by introducing certain assumptions about the parameters themselves based on experimental observations.

1. Firstly, the assumption was made that the association rate constant of the synaptic complex from the fully occupied excision substrate (k_3) is faster than the reassociation rate of the synapse from two recombined products (k_{-5}). This assumption was made based on the fact that for both FLP and Cre, intermolecular recombination is slower than excision (figure 4.17). However, the formation of both of these complexes very probably involves protein - protein interactions of the same nature, therefore the difference in the apparent rate constants governing their formation is due to the fact that in the excision substrate, the target sites are on the same molecule, thus limiting the volume in which two bound sites can diffuse. The distance between target sites in the FLP and Cre substrates is identical, (1.1kb) therefore if the reasonable assumption is made that synapsis occurs by random collision for both FLP and Cre (see introduction, section 1.3.2 for a discussion of this point) the factors governing the difference in k_3 and k_{-5} can be assumed to be identical for both FLP and Cre. More simply expressed, the assumption was made for the purposes of the model, that $k_3 = n.\ k_{-5}$, and that n(FLP) = n(Cre).

2. Secondly, the assumption was made that the dissociation rate constant of the unexcised synapse (k_{-3}) is the same as the dissociation rate constant of the excised synapse (k_5). This assumption was made based on the reasoning that the protein-

protein interactions holding the synapse together before and after recombination are similar. The assumption is not supported by any experimental evidence, and may be invalid, but was made in order to simplify the number of unknown parameters.

3. Finally the assumption was made that the back and forward rates of recombination within the synapse are identical. This assumption was made based on the observation that both FLP and Cre resolve synthetic Holliday junctions to give an equal mix of products (Dixon and Sadowski 1995, Sadowski 1995). This reasoning is not entirely validated by these experiments, which only test the direction of Holliday junction resolution and do not rule out the possibility that in a true reaction, one pair of strand exchanges occurs faster than the other. The fact that the chemistry of strand cleavage and ligation are identical for both pairs of strand exchanges (Jayaram et al, 1988), gives support to the idea that the back and forward rate of recombination may indeed be identical.

In further support of assumptions 2 and 3, is the fact that on an inversion substrate for both FLP and Cre, an equal mixture of both possible products is observed at equilibrium, and that on linear substrates containing two target sites with symmetrical spacers, excision and inversion occur *in vitro* with equal frequency for both FLP and Cre (Senecoff and Cox 1986, Hoess et al, 1986,).

The values of the parameters shown in Table 4.5 satisfy the three constraints described above, and give the best fit of the model to the time course data shown in figure 4.14 e,f) for FLP and Cre. The recombination time course data simulated by the model is shown superimposed on the measured data points in figures 4.19 a and b. A near perfect fit was obtainable for both proteins, using the DNA binding parameters measured (Table 4.3 and 4.5) and the protein and substrate inputs as used in the experiments in figure 4.14e, f.

For the unknown parameters, some of the values shown in Table 4.5 are unique: variations of two fold or more in either direction cause the simulated curve to deviate significantly from the data. For example, the pair of parameters k_{-3} and k_5 describing

Table 4.5 Parameters used for fitting simulated recombination time course

Rate constant	Description	FLP	Cre
k_1 (M^{-1} s^{-1})	association of first monomer	1.0×10^{7}	2.2×10^{8}
k_{-1} (s^{-1})	dissociation of first monomer	1.0×10^{-1}	2.4×10^{-2}
k_2 (M^{-1} s^{-1})	association of second monomer	4.0×10^{6}	3.0×10^{8}
k_{-2} (s^{-1})	dissociation of second monomer	1.3×10^{-2}	1.8×10^{-3}
k_3 (s^{-1})	synapse formation	8.3×10^{-2}	2.5
k_{-3} (s^{-1})	synapse dissociation	1.2×10^{-2}	6.6×10^{-3}
k_4 (s^{-1})	recombination forward	1.6×10^{-1}	1.6×10^{-1}
k_{-4} (s^{-1})	recombination backward	1.6×10^{-1}	1.6×10^{-1}
k_5 (s^{-1})	recombined synapse dissociation	1.2×10^{-2}	6.6×10^{-3}
k_{-5} (M^{-1} s^{-1})	recombined synapse reassociation	1.6×10^{5}	3.3×10^{7}

Figure 4.19 Simulated time course curves for FLP and Cre recombination

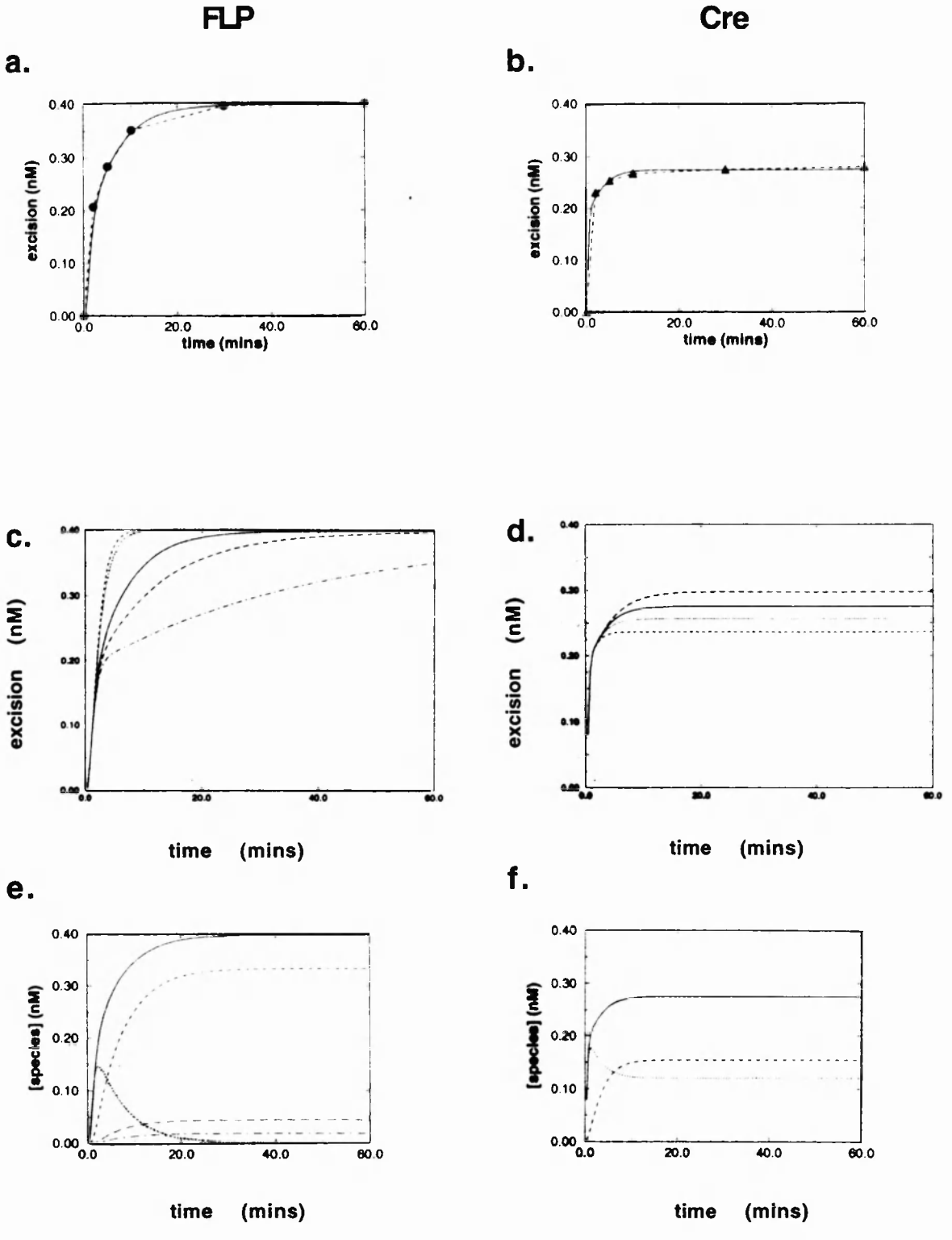
The parameters shown in Table 4.5 were used to simulate time course curves representing total recombination by FLP or Cre. **a)** FLP. *Solid line:* simulated time course. *Filled circles, dashed line*; the data points from figure 4.14e are superimposed on the plot. **b)** Cre. *Solid line:* simulated time course. *Filled triangles, dashed line*; the data points from figure 4.14f are superimposed on the plot.

c) Effect of varying both of the parameters k-3 and k5 on the simulated recombination time course for FLP. *Dotted line*, 5 fold increase; *dashed line,* 10 fold increase; *long dashed line,* 2 fold decrease; *dot dashed line,* 7 fold decrease.

d) Effect of varying the parameter k-5 on the simulated recombination time course for Cre. *Dotted line,* 2 fold increase; *dashed line*, 5 fold increase; *long dashed line*, 2 fold decrease.

e) f) The mathematical model describing recombination over time gives not only the concentration of final recombination product, but also the concentration of all intermediate species. In figures c) and d), the simulated time course curves from a) and b) above are shown as *solid lines.* Other intermediate species are shown on both plots as follows: *Dotted line*, excision product in synaptic complex (IEP). *Dashed line*, excision product with two monomers bound (EPM2). *Long dashed line*, excision product with one monomer bound (EPM). *Dot dashed line,* free excision product (EP). In the model, the total excision product (EPtot) = IEP + EPM2 + EPM + EP. Note that for Cre, (fig d) no EPM or EP are seen.

Figure 4.19


FLP
Cre
a.
b.
excision (nM)
time (mins)
0.00
0.10
0.20
0.30
0.40
0.0
20.0
40.0
60.0
c.
d.
excision (nM)
time (mins)
e.
f.
[species] (nM)
time (mins)

dissociation of the synapse before and after recombination respectively are unique for both proteins. The effect of varying this pair of parameters on the fitted FLP curve is shown in figure 4.19c. The parameter k_{-5}, describing the reassociation of protein bound recombined products after recombination, is also unique to the fits shown. The effect of varying this parameter on the fitted Cre curve is shown in figure 4.19 d. Figures 4.19c and d illustrate that the nature of variation in the simulated curve is different for different parameters. Some variations (e.g., figure 4.19c, long dashed line and 4.19d, dotted line) fall within the type of experimental variations which have been observed. Other deviations, for instance all other lines in figure 4.19c, have not been observed experimentally. This allows upper and lower limits to be set for the values of the parameters k_{-3}, k_5 and k_{-5} (i.e 2 fold greater or smaller than the values given in Table 4.5)

Regarding the other parameters, the values shown in Table 4.5 are not unique: large variations in one or both directions have no effect on the fit of the curve to the data. For instance the values of k_4 and k_{-4} describing the back and forward rate of recombination within the synapse, shown in table 4.5 are lower limits. A two fold reduction in k_4/k_{-4} causes a slight reduction in the initial rate of simulated recombination for both FLP and Cre, but has no effect on the total recombination, whilst a ten fold reduction reduces initial recombination to such an extent that the total recombination at 60 minutes is also reduced. However increases of up to 1000 fold in k_4/k_{-4} has no effect at all on the shape of the simulated curve. Thus for this parameter, a lower limit, but not an upper limit can be set for both proteins.

The values of parameter k_3, describing the formation of the synaptic complex from the species SM4, are only unique if the condition that $k_3 = n.\ k_{-5}$, and that n(FLP) = n(Cre) (assumption 1 above) is satisfied. If this condition is abandoned, then the lower limit of k_3 for both FLP and Cre is the same ($8.3 \times 10^{-2}\ s^{-1}$). Below this limit the initial rate of recombination is reduced for both FLP and Cre. The upper limit for Cre is $2.5\ s^{-1}$. Values above this limit cause an increase in the total level of recombination. An upper limit of k_3 for FLP was not found.

In summary, these results show that the mathematical model can simulate the time course data observed for both FLP and Cre. The values of many parameters are unique to the fit obtained, whereas for other parameters, broad limits can be defined.

4.3.7 Parameters responsible for the different behaviour of FLP and Cre: Models for recombination

One of the most striking differences in FLP and Cre recombination kinetics observed experimentally is the difference in maximum recombination levels achieved by both proteins. Cre reaches 50 to 75%, whereas FLP can recombine 100 % of its substrate. This behaviour is reproducible over a range of protein: substrate ratios in each case (see for instance, figure 4.14a, b).

Mathematical modelling of the time course of recombination shows that the most important parameter for determining this behaviour is k_{-5}, describing the reassociation of the excised synaptic complex. As shown in the preceding section, the values of this parameter can be uniquely determined for FLP and Cre (1.6×10^5 and 3.3×10^7 M^{-1} s^{-1} respectively). A theoretical value for the equilibrium constant K_5, describing the stability of the excised synapse, is given by the relationship k_{-5}/k_5 (note k_5 can also be determined uniquely). Conversion of the theoretical K_5 values to kcal/mol for FLP and Cre at their respective incubation temperatures gives 9.83 and 13.94 kcal/mol for FLP and Cre respectively, a difference in free energy of 4.11 kcal/mol. The predicted effect of this theoretical difference in stability of the FLP and Cre synaptic complexes is illustrated in figures 4.19e and f, in which the simulated excised synaptic complex is shown as a dotted line. For FLP, this species is short lived, decaying to zero after 30 minutes (dotted line, figure 4.19 e). For Cre, however, the excised synaptic complex (dotted line, figure 4.19f) remains present throughout the reaction.

In summary, the mathematical modelling experiment predicts that the most important parameter for determining the equilibrium

level of recombination is the stability of the synaptic complex after recombination. The model predicts that this complex is more stable for Cre than for FLP, and that this is the reason for the lower maximum level of recombination observed for Cre. Bearing in mind that the rate of synapsis for unrecombined and recombined molecules are related, this prediction is consistent with the result of the cointegration assay, (figure 4.18) which suggests that the rate of synapsis for Cre is indeed faster than for FLP.

Theoretical values for the rate constants and equilibrium constants governing the stability of this complex are suggested by the model. The behaviour of the complex over time is predicted by the model to be very different for FLP and for Cre. This analysis thus offers a concrete set of predictions which could be experimentally tested if an assay for synapsis suitable for kinetic analysis were available.

What is the contribution of the different DNA binding affinities of FLP and Cre to their behaviour? The main effect of the parameters k_1, k_{-1}, k_2 and k_{-2} on the mathematical model is to determine the amount of input protein required for a given amount of recombination. Thus, in the model, a lower DNA binding affinity means that a higher protein concentration is required to produce a recombination event. Using the parameters in Table 4.5, and varying the protein input in the model, the protein titration data for FLP and Cre shown in figure 4.14c and d can be simulated reasonably accurately (the protein input in the model is within 1.5 fold of the observed protein input for any given amount of recombination for FLP and Cre, not shown).

It has been suggested that the requirement for stoichiometric rather than enzymatic amounts of FLP and Cre may reflect non enzymatic behaviour, or inactivity of protein (Gates and Cox ,1988, Abremski and Hoess, 1984). My results show clearly that the requirement for stoichiometric amounts of both proteins is entirely consistent with enzymatic behaviour (i.e. turnover, see equations 28 to 42) and with 100% activity of both protein preparations. The expectation that enzymatic quantities of FLP and Cre should be required for catalysis, is not applicable to these reactions because they are not of the "one enzyme -one substrate" class. The fact that

the excision reaction involves interactions between four enzyme molecules, and two substrate sites is sufficient to explain the requirement for stoichiometry. The full occupation of all four sites is necessary before a recombination event can take place.

The difference in DNA binding affinities of FLP and Cre also has an effect on the distribution of the different bound species in the simulated recombination reaction. In Figure 4.19 e and f, excision products with two, one or no monomers bound are shown. For FLP, (figure 4.19e) the synaptic complex (dotted line) decays rapidly into bound excision products (dashed lines). The species EPM2, with two monomers bound (dashed line), is formed first, and then gives rise to a small amount of species EPM (long dashed line) with one monomer bound. This monomer dissociates relatively readily to give the species EP (dot dashed line) which is the free excision product. In contrast, for Cre, (figure 4.19f), only the bound species EPM2 (dashed line), with two monomers bound, is detected. This species is formed at the expense of the excised synaptic complex, IEP (dotted line) and an equilibrium is reached between IEP and EPM2 within ten minutes. No formation of the excision product- monomer complex or the free excision product is seen, a result of the high affinity of the Cre dimer complex for the LoxP site. The total recombination plotted by the model (solid line on all plots in figure 4.19), is the sum of the DNA in the excised synapse, and the bound and unbound forms of the excision product.

4.3.8 Simulation of substrate titration experiments for FLP and Cre. Substrate saturation behaviour is determined by protein consumption.

The mathematical model accurately describes the recombination behaviour of FLP and Cre, enabling simulation of both time course and protein titration experiments with a single set of parameters for each protein. To ask whether the model would also behave realistically in the substrate titration experiment, time course curves were simulated at different substrate inputs, using the parameters in Table 4.5 for FLP and Cre. The protein input was held constant, as for the real substrate titration experiment (figure

4.15) at 25.6nM FLP, 3.2 nM Cre. In figure 4.20 a and b, the result of the simulated substrate titration experiment is shown. The nM total excision product at 2 minutes and at 60 minutes is plotted against substrate concentration. Comparison of the simulated experiment with the real experiment (figure 4.15c and d) shows clearly that the model qualitatively reproduces the behaviour of FLP and Cre. In both the real and the simulated experiment, FLP recombination at 60 minutes increases linearly over substrate concentration, whilst the initial rate drops off gently. Interestingly, at the highest substrate concentration (20nM) a reduction in total FLP recombination is predicted by the model. For Cre, in both the real and simulated experiments, the initial and final recombination levels increase sharply with substrate concentration up to 0.5 to 1nM substrate, after which point both curves descend rapidly with increasing substrate concentration. The protein monomer to half site ratio represented by the maximum level of initial and final recombination for Cre in both the simulated and the real experiments is between 1: 1 and 2:1.

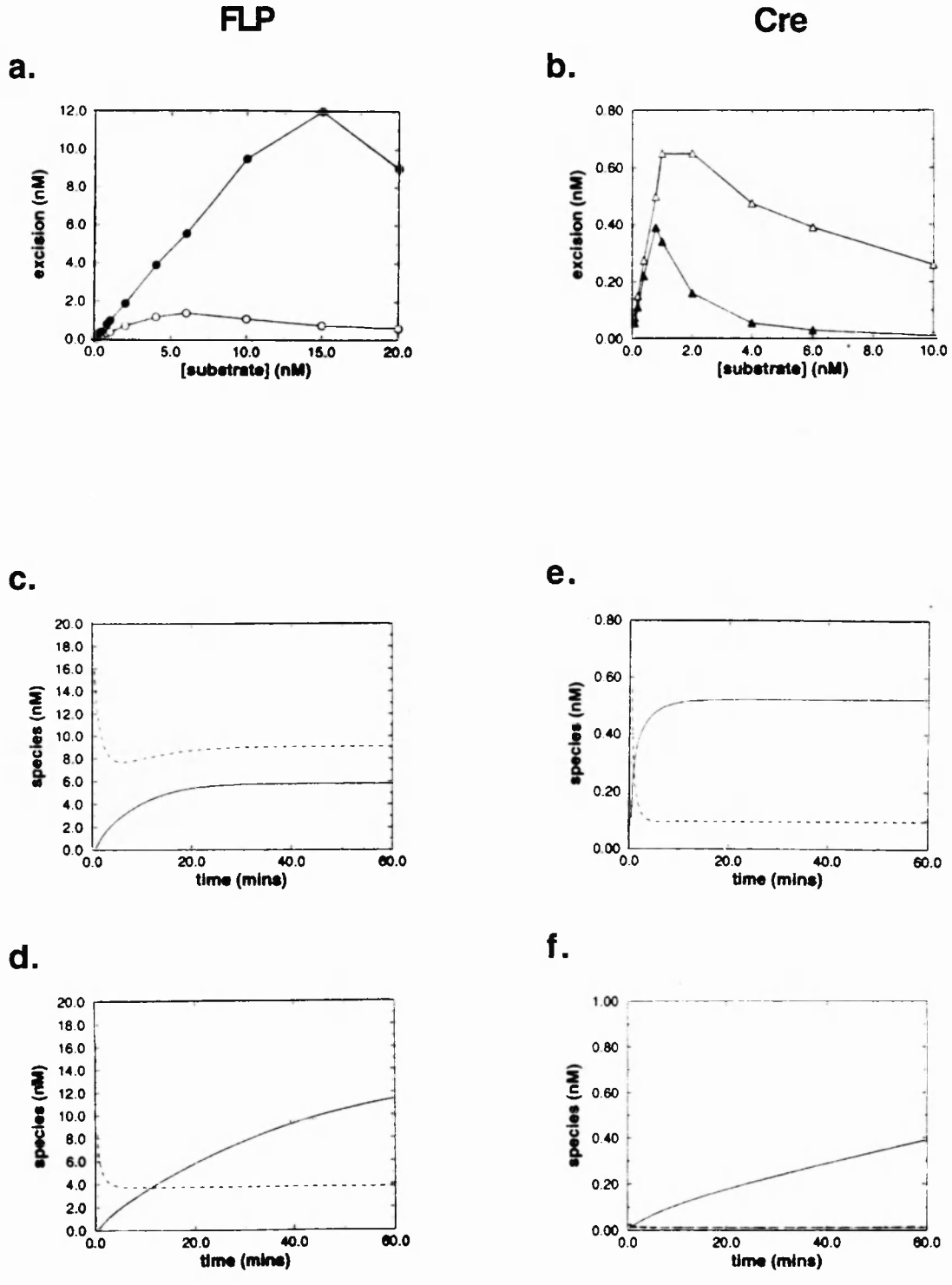
Although the simulated data qualitatively reproduces the observed behaviour of FLP and Cre, the quantitative fit is only approximate. Generally, higher levels of recombination are predicted by the model than are observed experimentally (note however, that for Cre the total percent simulated recombination does not exceed 65% for any given substrate concentration). Such variations may be due to experimental variation between the experiments used to fit the model (figure 4.14) and the substrate titration experiments (figure 4.15). These experiments were performed on different days with different substrate preparations, and were also processed differently. (The substrate in figure 4.14 was directly labelled, whilst figure 4.15 was quantified by Southern hybridisation). Several repeated experiments would be necessary to determine the extent of experimental variation. The effect of varying different parameters on the simulated substrate titration curves as described in section 4.3.6 was assessed, with essentially the same result; each parameter affected the simulated time course in the same way at all substrate concentrations tested (data not shown). Within the limits defined (section 4.3.6), there was no parameter or

Figure 4.20 Simulated substrate titration curves for FLP and Cre recombination

Using the input values from Table 4.4 for FLP and Cre, time course curves were simulated using the following substrate input concentrations: 0.1, 0.2, 0.4, 0.8, 1, 2, 4, 6, 10, 15 and 20 nM. The nM concentration of excision product at 2 minutes and at 60 minutes for each substrate input was plotted against substrate concentration. **a)** FLP. *Open circles*, nM excision at 2 minutes. *Closed circles*, nM excision at 60 minutes. **b)** Cre. *Open triangles*, nM excision at 2 minutes. *Closed triangles*, nM excision at 60 minutes. **c), d).** Simulated time course curves for FLP recombination at 6 nM (c) and 15 nM (d) substrate input, showing the nM concentration of total recombination product (*solid lines*) and of free protein monomer (*dashed lines*) plotted against time.

e), f). Simulated time course curves for Cre recombination at 0.8 nM (e) and 6 nM (f) substrate input, showing the nM concentration of total recombination product (*solid lines*) and of free protein monomer (*dashed lines*) plotted against time.

Figure 4.20


FLP
Cre
a.
b.
excision (nM)
[substrate] (nM)
c.
e.
species (nM)
time (mins)
d.
f.

combination of parameters whose variation was sufficient to account for the differences observed in the simulated and real data. Thus, if the deviation from the observed data is not entirely accountable for by experimental variation, this suggests either that further parameters are at work at higher DNA concentrations (such as, for example, non specific DNA binding), or that one or more of the three assumptions used to constrain the parameters are invalid. Since the deviation of the simulated from the real data is small (around two fold or less), and the qualitative behaviour is reproduced faithfully, the introduction of further parameters and the abandonment of assumptions is not pursued further here.

To examine the possible reasons for the substrate saturation behaviour of FLP and Cre, the mathematical model was used to simulate time course curves at optimal and suboptimal substrate concentrations, enabling prediction of the concentration of free protein monomer over the course of the reaction (figure 4.20 c to f). Figure 4.20a shows that FLP reaches maximum initial recombination at 6nM substrate. In figure 4.20 c, the total recombination (solid line) and the free protein monomer generated by the model are plotted against time. The concentration of free monomer remains above 8 nM throughout the reaction, meaning that only two thirds of the total input protein are involved at any one time in the reaction. This prediction is consistent with the observation of Waite and Cox (1995). Using a similar FLP: FRT ratio, the authors showed that when free substrate is added to a recombination reaction at equilibrium, the newly addded substrate is substantially recombined, suggesting that free FLP monomers are present in the reaction.

Simulation of recombination products and free protein concentration for FLP at 15 nM substrate shows that more protein is consumed by the reaction in this case (figure 4.20d). For Cre, at the optimal substrate concentration of 0.8 nM, the simulation of free protein (figure 4.20e, dashed line) shows that most of the input protein is consumed by the reaction. Less than 3 % of the input protein is free. At a higher substrate concentration, 6nM, at which initial and final recombination are reduced, figure 4.20f shows that the free protein concentration is greatly reduced (dotted line).

In summary, this analysis suggests that the main reason for the rapid substrate saturation observed with Cre is an exhaustion of the supply of free monomer. This prediction is supported experimentally by the observation that the point of Cre saturation can be shifted to a higher substrate concentration by increasing the Cre protein input in the substrate titration experiment (not shown). For FLP, the model predicts that the ability to recombine higher concentrations of substrate is due to the fact that more free protein is present. In the model, these characteristics of FLP and Cre are determined by their DNA binding affinities which in turn define how much active protein is required to recombine a given amount of substrate. These results suggest that the behaviour of a protein in the substrate titration assay can give an indication both of the true concentration of active monomer, and of its DNA affinity.

The results presented here also show that the mathematical model is useful both for investigating possible reasons for the differences between the two proteins, and also for generating concrete, quantitative predictions, which can be tested experimentally.

4.4 Discussion

In this chapter, I have presented a comparative kinetic analysis of FLP and Cre, in which rate constants for various steps of the recombination reaction are determined directly, or suggested by mathematical analysis. In the first part of the chapter, the DNA association and dissociation rates for both the first and second monomer binding to the inverted repeat target site were determined directly for FLP and Cre. This analysis showed that both FLP and Cre bind their target sites in a cooperative manner, and that for both proteins, the increased affinity observed upon binding of the second monomer is mostly due to a slow dimer dissociation rate. This analysis also showed that Cre has a higher affinity for the LoxP site than FLP for the FRT, and that this higher affinity is due both to a faster association rate and to a slower dissociation rate, for both the first and the second monomer. The DNA binding affinities of FLP and Cre have not previously been determined.

Crucial to the analysis presented in this chapter is the activity of the protein preparations used. I have shown that the Cre preparation is close to 100% active, that the FLP preparation is at least 56% active. Furthermore, using the substrate titration assay, I have shown that this is almost certainly an underestimation of the true fraction of active FLP protein. This analysis shows that stoichiometric quantities of active protein monomer are required for efficient recombination, and indicates that this requirement is a result of the cooperative nature of the reaction.

The results presented in the second part of this chapter show that FLP can recombine 100% of its target, whilst Cre recombines only 70%. Mathematical analysis suggests that for Cre this reflects an equilibrium between excision and reintegration, and shows that the most likely reason for this behaviour is a slow synaptic dissociation rate.

The recombination characteristics of *in vitro* translated proteins were found to be similar to those of purified preparations, enabling comparison of the Kw protein with FLP and Cre, suggesting that the DNA binding affinity of Kw is similar to that of FLP.

The difference in DNA binding affinities measured for FLP and Cre is sufficient to account for the fact that more FLP is required than Cre for a given amount of recombination, and may have implications for the design of *in vivo* experiments. Although caution must be exercised in evaluating the relevance of *in vitro* analysis to the situation *in vivo,* certain possibilities are raised by these results. Firstly, for FLP the expression of high levels of recombinase may be necessary in applications where efficiency is required. Secondly, in applications where recombinase repression is required, lower background levels of Cre than of FLP may be sufficient to produce an unwanted recombination event.

A further possibility raised by the difference in DNA binding affinities of FLP and Cre concerns the role of the third direct repeat element present in the wild type FRT. The presence of this repeat element enhances binding of the monomers at all three sites in a

DNAse footprint assay (Andrews et al, 1987, Beatty and Sadowski, 1988) but apparently has no effect on the efficiency of recombination *in vitro.* The results presented in this chapter show that the low DNA binding affinity of FLP for the minimal target site only limits the rate and extent of recombination at limiting protein concentration, raising the possibility that for the experiments in which the third repeat element was found to have no effect on *in vitro* recombination rates, the concentration of FLP was not limiting. The third repeat element stimulates intermolecular recombination in *E. coli,* (Jayaram, 1985), suggesting that in this assay, where synapsis would be expected to occur slowly, the DNA binding affinity is important. To further investigate this point, it would be interesting to compare the recombination and DNA binding kinetics of FLP at limiting protein concentration on substrates with and without the third repeat element.

A further speculation which arises from the idea that the third repeat element may increase the affinity of FLP for the FRT, is that in other yeast 2 micron-like plasmids, the number of additional repeats may give an indication of the affinity of the recombinase for a single repeat element. This idea is based on pure speculation, but is consistent with the observation that for R recombinase from *Zygosaccharomyces rouxii,* whose wild type target site contains four extra repeat elements, all of which are bound by recombinase (Araki et al., 1992) no bandshift is detectable when the inverted repeat target alone is used (Lee et al., 1992). This idea could be tested experimentally by comparison of the recombination efficiencies of different recombinases at limiting protein concentrations on target sites with and without extra repeats. If the extra repeats are important for increasing the DNA binding affinity of yeast family recombinases, then improved substrates for *in vivo* experiments could be envisaged, in which one or many extra direct repeat elements are included in the target site.

The mathematical model for recombination has proved useful for examining the parameters which may be responsible for dictating the very different behaviours of FLP and Cre in the *in vitro* excision assay. In particular, the model predicts that the stability of the synaptic complex is higher for Cre than for FLP. The usefulness of

such a model is illustrated by the fact that for the synaptic complex, concrete predictions can be made, regarding both its survival during the recombination reaction, and the free energy of association dictating its stability. For FLP, the transient nature of the synaptic complex predicted by the model may account for the lack of success in isolating such an intermediate (Sadowski, 1995). However it is possible to trap an intermediate that has the properties of a FLP synaptosome using the protein cross linking agent, glutaraldehyde (Amin et al., 1990, 1991). It is possible that such an assay could be adapted for quantitative kinetic analysis, to test the predicted differences in the stability of the synaptic complex for FLP and Cre.

Waite and Cox (1995) suggest an alternative model for FLP synapse dissociation, in which one or more monomers may first dissociate from the synapse, rendering it incompetent for recombination. The dissociation of the remaining three monomers is suggested to be slow, with the result that substrates are sequestered and unavailable for recombination. This model is not inconsistent with the mathematical predictions made here. The "dissociation" of the synaptic complex predicted by the mathematical model, may in fact correspond to the dissociation of the first monomer suggested by Waite and Cox. My experimental data do not give information about susbtrate turnover. It may be interesting to incorporate the mechanism suggested by Waite and Cox into the mathematical model described here, to investigate the predictions made by such a model for recombination under different circumstances.

The measured differences in DNA binding affinities for FLP and Cre, in combination with the predicted differences in synaptic stability, may have further implications for their applied use *in vivo*. In particular, these parameters might affect the efficiency of site specific integration, an application which is particularly interesting for the purposes of gene targeting in organisms such as plants and flies, where homologous recombination is not easily achieved. Site specific integration is inefficient, both because the initial integration event itself is rare, and because the integrated DNA is efficiently reexcised. For a recombinase with higher DNA binding affinity, and a faster rate of synapsis, the initial site

specific integration event may be expected to occur more efficiently. However, the higher the DNA binding affinity, and the more stable the synaptic complex, the higher is the probability that the integrated DNA will be reexcised. This reasoning suggests that there may be an ideal combination of DNA affinity and synaptic stability for the purposes of site specific integration, which represents the best compromise between integration efficiency and excision inefficiency. This "ideal compromise" of characteristics may not be possessed either by FLP or by Cre, but may be reached by on the one hand, manipulation of DNA affinity by target site design, and on the other, manipulation of synaptic stability by protein mutation.

CHAPTER 5
Different thermostabilities of FLP and Cre recombinases

5.1 Introduction

Both FLP and Cre recombinases have been used for the genomic manipulation of a wide range of different organisms, including bacteria, yeast, plants, flies and mammals (for review, see Kilby et al, 1993). Although much work has focused on the reaction mechanisms of the two recombinases, little has been done to define the characteristics which may be important for their applied use in living systems. In the preceding chapter, I compared various kinetic properties of FLP and Cre, including DNA binding affinity and rate of synapsis, and discussed how these parameters may affect their behaviour in living systems.

In our lab we are also investigating other properties of FLP and Cre which might be important for their optimal use *in vivo*, including the effect of temperature (Buchholz et al, 1996). In this chapter, I show that FLP recombinase is thermolabile in an *in vitro* recombination assay, whereas Cre is thermostable, and that this difference occurs at the level of DNA binding. In addition, two FLP mutants are described, one of which has increased temperature sensitivity, and the other of which is more thermostable than wild type FLP.

These results suggest that the growth temperature of the host organism is an important factor in determining the efficiency of recombination. The results are discussed in the context of optimal use of FLP and Cre for genomic manipulation experiments.

5.2 Results

5.2.1 FLP is a temperature sensitive recombinase *in vitro*

Upon comparison of the properties of FLP and Cre recombinases in *E. coli,* the observation was made by others in our lab that FLP but not Cre, is sensitive to temperature. (Buchholz et al., 1996). Both recombinases were fully active between 34 and 39°C in this assay. Cre was fully active at temperatures of up to 46°C, whereas FLP activity was inhibited above 39°C.

To test whether this temperature effect could be observed *in vitro*, I used purified FLP and Cre proteins and the linearised excision substrates pSVpaz and pSVpax described in chapter 4 (section 4.3.1). *In vitro* excision assays were performed at different temperatures in time course experiments using linearised substrates (figure 5.1). The experiment was carried out for FLP and Cre at two different protein concentrations. In figures 5.1a and 5.1c, the FLP or Cre concentration used was determined from protein titration experiments (see for example, chapter 4, section 4.3.1, figure 4.14) as the minimum amount of protein required to give maximum recombination of 0.4nM substrate in 60 mins. For FLP at this protein concentration, recombination is most efficient at 23 and 30°C. Recombination efficiency is reduced with increasing temperature (figure 5.1a, e).

In the analogous experiment for Cre, (figure 5.1c), no such effect of temperature is observed. Cre recombines its target equally well at all the temperatures tested, reaching its characteristic maximum of 60% excision at all temperatures tested.

For FLP at this protein input, (figure 5.1a) the reduction in recombination efficiencies at higher temperatures does not appear to be due to an overall reduction in the reaction rate. The amount of recombination at 5 minutes is the same at all temperatures except for 42°C, showing that there is no reduction in initial protein activity at temperatures of 39°C and below. This suggests that the reduced activity at higher temperatures may be due to an effect on protein stability, i.e, that the active conformation of the protein has a shorter half life at higher temperatures. This hypothesis predicts that if the initial rate of the reaction were increased by increasing the protein concentration, then a higher level of recombination could be reached before the protein is inactivated.

Figure 5.1b shows that this is indeed the case. A two fold increase in the FLP protein input is sufficient to increase the initial rate of the excision reaction substantially (compare the 2 minute time points in figure 5.1a and 5.1b at all temperatures). Under these conditions, FLP is able to recombine 70 to 80% of its substrate at

Figure 5.1 Effect of temperature on recombination by FLP and Cre.

See chapter 3, figure 3.2a (Kw in vitro recombination assay) for description of assay.

a) to d) Different amounts of purified FLP and Cre proteins were incubated with 0.4nM of the appropriate excision substrate at each of the temperatures shown, and aliquots removed at indicated time points. Unrecombined (4.0kb) and recombined (2.9kb) bands are shown on the right of the figure. Protein concentrations in a) and d) ("1X") were determined from figure 4.17c and d as the minimum amount of protein required to give maximum recombination of 0.4nM substrate in 60 mins at 30°C (FLP) or 37°C (Cre).

a) FLP 1X. 25.6nM FLP (protein monomer:1/2 site ratio =16:1).

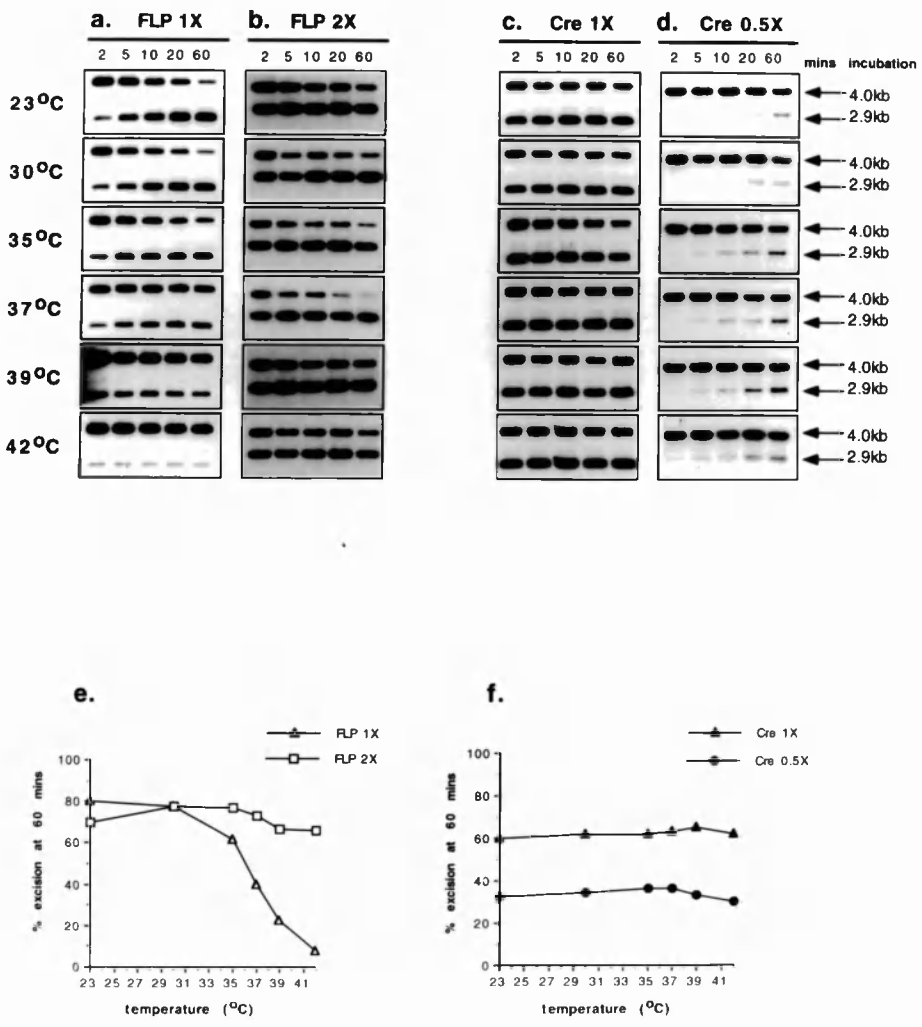
b) FLP 2X. 51.2nM FLP (protein monomer:1/2 site ratio =32:1).

c) Cre 1X. 3.2nM Cre (protein monomer:1/2 site ratio =2:1).

d) Cre 0.5X. 1.6nM Cre (protein monomer:1/2 site ratio =1:1).

e), f) % excision after 60 mins in a) to d) is plotted against temperature at the FLP and Cre concentrations as indicated. **e)** FLP *Open triangles*, FLP 1X; *open squares*, FLP 2X. **f)** Cre. *Closed triangles*, Cre 1X; *closed circles;* Cre 0.5 X.

Figure 5.1


a. FLP 1X
b. FLP 2X
c. Cre 1X
d. Cre 0.5X
2 5 10 20 60
mins incubation
23°C
30°C
35°C
37°C
39°C
42°C
4.0kb
2.9kb
e.
FLP 1X
FLP 2X
% excision at 60 mins
temperature (°C)
f.
Cre 1X
Cre 0.5X
% excision at 60 mins
temperature (°C)

all temperatures. Thus the reduction in FLP activity at higher temperatures can be overcome by increasing the protein concentration.

This led to the question of whether Cre would behave in a temperature sensitive manner if the initial rate of recombination were reduced by reducing the protein input. To test this idea, I carried out the experiment shown in figure 5.1d. The concentration of Cre was reduced two fold, and the time course experiment was repeated. Reducing the Cre concentration severely reduced the initial rate of recombination (compare the 2 minute time point in figures 5.1c and 5.1d at all temperatures), and the maximum recombination achieved in 60 minutes. The initial rate and maximum level of recombination were however unaffected by the incubation temperature. This result shows that the lack of temperature sensitivity shown by Cre at optimum protein concentration is not merely due to the speed with which it recombines its substrate.

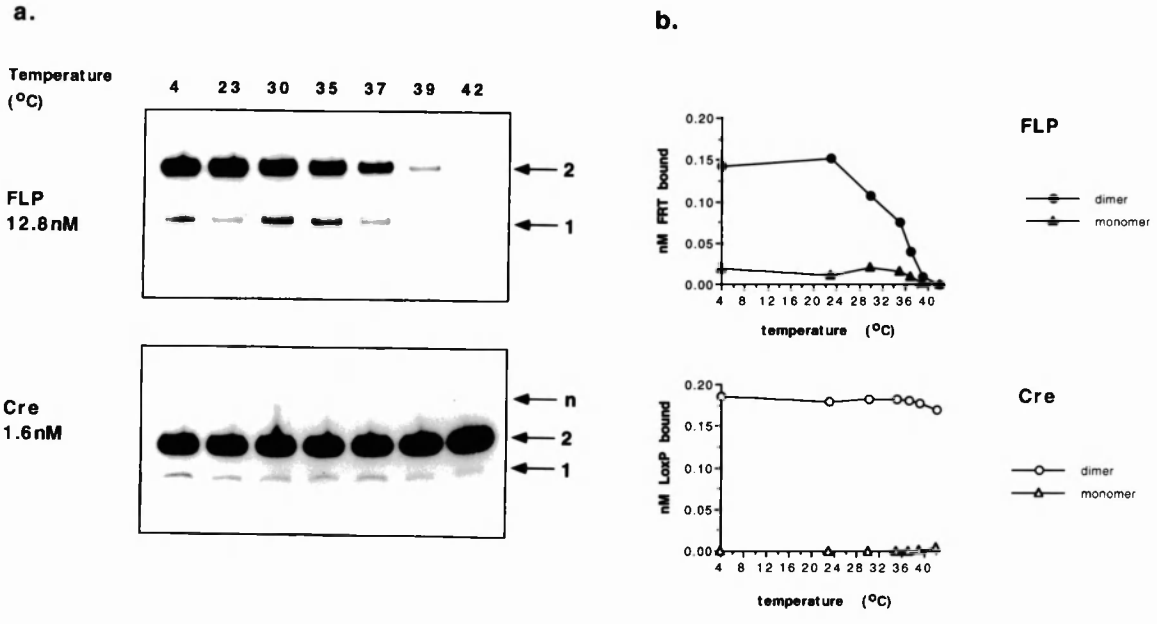
5.2.2 Temperature sensitivity occurs at the level of DNA binding

Using a bandshift assay, and wild type FLP which was 50% pure Friesen and Sadowski (1992) showed that FLP binds its substrate equally efficiently at 30 and 37°C, but that binding is severely reduced at 42°C. Since the incubation times in these experiments were long (30 minutes at 30°C plus 30 minutes at the higher temperature) relative to the times at which I have observed loss of activity at temperatures above 30°C in the recombination assay, (5 to 10 minutes), I investigated the DNA binding properties of FLP and Cre under the conditions used in the previous section for the recombination assay. FLP or Cre were incubated for ten minutes at different temperatures with 0.2nM of the appropriate "full site" bandshift substrate, carrying the inverted repeat element which consists of two recombinase binding sites (see chapter 4, section 4.2.1 and figure 4.2 for a full description of these substrates). The protein concentrations used are defined from protein titration

Figure 5.2 Effect of temperature on DNA binding by FLP and Cre.

a) 12.8nM purified FLP or 1.6nM purified Cre protein was incubated for 10 minutes with 0.2nM of the appropriate full site bandshift substrate (see chapter 4, figure 4.2 for full description of substrates). **b)** The bands in (a) were quantified by PhosphorImager analysis and nM DNA bound plotted against temperature. *Closed circles*, FLP dimer; *closed triangles*; FLP monomer; *open circles*, Cre dimer; *open triangles*; Cre monomer.

Figure 5.2


a.
Temperature (°C)
4 23 30 35 37 39 42
FLP 12.8nM
2
1
Cre 1.6nM
n
2
1
b.
nM FRT bound
0.20
0.15
0.10
0.05
0.00
4 8 12 16 20 24 28 32 36 40
temperature (°C)
FLP
dimer
monomer
nM LoxP bound
0.20
0.15
0.10
0.05
0.00
4 8 12 16 20 24 28 32 36 40
temperature (°C)
Cre
dimer
monomer

recombination assays as the minimum amount sufficient to give maximum recombination of 0.2nM excision substrate in 60 minutes. (Note that this corresponds to the highest concentration of each protein used for the kinetic analysis of chapter 4 , section 4.2).

In the bandshift assay, FLP and Cre each form two complexes on their substrate, corresponding to occupation by a recombinase monomer of one or both halves of the site. In this assay, FLP DNA binding is most efficient at 23°C and below (figure 5.2a, b) Binding is slightly reduced at 30°C and increasingly reduced at higher temperatures. Interestingly, at 30 and 35°C, although the amount of dimer complex is reduced, the monomer complex is increased, suggesting that the temperature induced reduction in the affinity of FLP for its DNA target may affect the dimer and the monomer complex differentially. The extent of Cre binding, and the relative distribution of monomer and dimer complex, is identical at all the temperatures tested. This result is consistent with the results from the *in vitro* recombination assay at equivalent protein inputs, and confirms that under these conditions, the temperature sensitivity of FLP in the recombination assay occurs at the level of DNA binding.

The result of Friesen and Sadowski (1992), showing that FLP binds its substrate equally efficiently at 30 and 37°C, may differ from the results presented here for several reasons. Firstly, the protein preparation used was only 50% pure. Several authors report that FLP becomes increasingly unstable with increasing purity (Meyer-Leon et al., 1984, Babineau et al., 1987). Secondly the substrate used by Friesen and Sadowski was the wild type target containing the third, direct repeat element in addition to the inverted repeat element. Although this third repeat element apparently has no effect on recombination efficiency *in vitro*, (Senecoff et al. 1985) footprinting analysis suggests that FLP binding to the third element stabilises the interaction at all three sites (Andrews et al., 1987). However if the stability of binding observed by Friesen and Sadowski were mediated by the third repeat element, then only the binding of complex 3 and not that of complexes 1 and 2 should be stabilised. This was not the case, as the authors showed that there was no difference in the relative distribution of the three

complexes at 37 and 30°C. The third possibility is that the protein concentration used by Friesen and Sadowski was sufficient to overcome the effect of temperature on DNA binding (the protein concentration used for the temperature experiment was not stated).

5.2.3 FLP mutants with altered temperature sensitivity.

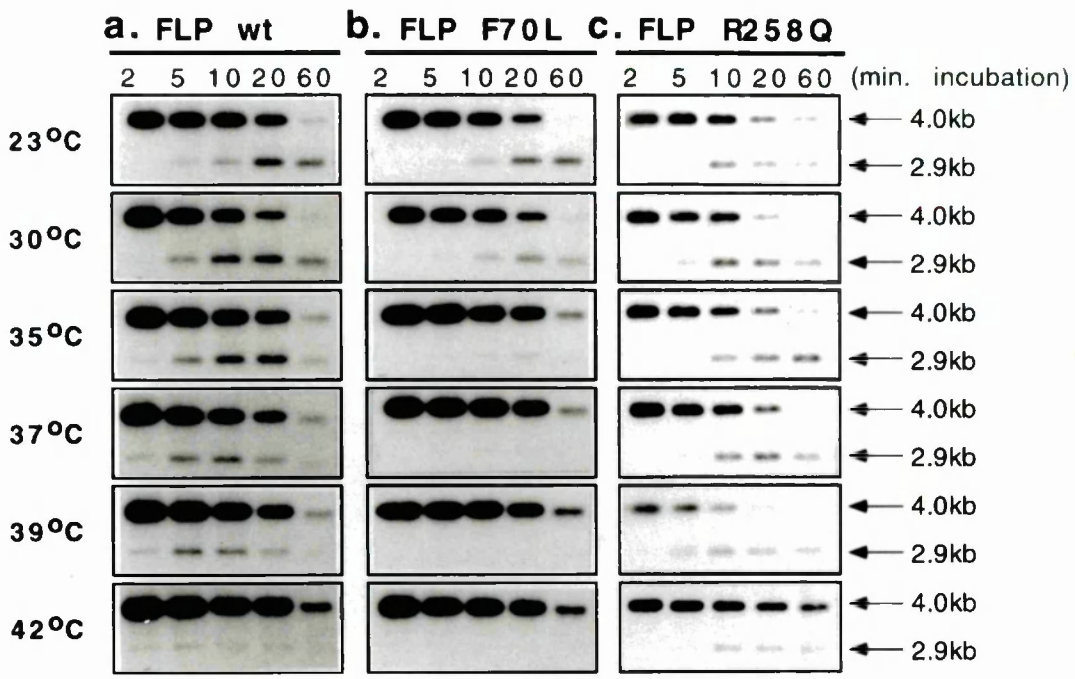
In the course of our experiments with FLP, it was found that the commercially available pOG44 plasmid (Strategene, O´Gorman et al., 1991)) carries a point mutation in the FLP coding region. This mutation changes amino acid 70 from Phe to Leu. Because purified mutant FLP protein was not available, the recombination activity of FLP F70L and wild type proteins was compared using *in vitro* translated proteins (figure 5.3a, b). Time course experiments were carried out at different temperatures. The protein input was defined from protein titration experiments as the minimum amount of protein required for maximum recombination of 0.04nM excision substrate in 60 minutes at 30°C (see chapter 4, section 4.3.3, figure 4.16). In this assay, FLP F70L recombines its substrate as efficiently as wild type FLP at 23 and 30°C, but significantly less efficiently than wild type FLP at 35, 37 and 39°C (figure 5.3 a, b, d). Thus F70L has the characteristics of a temperature sensitive mutation.

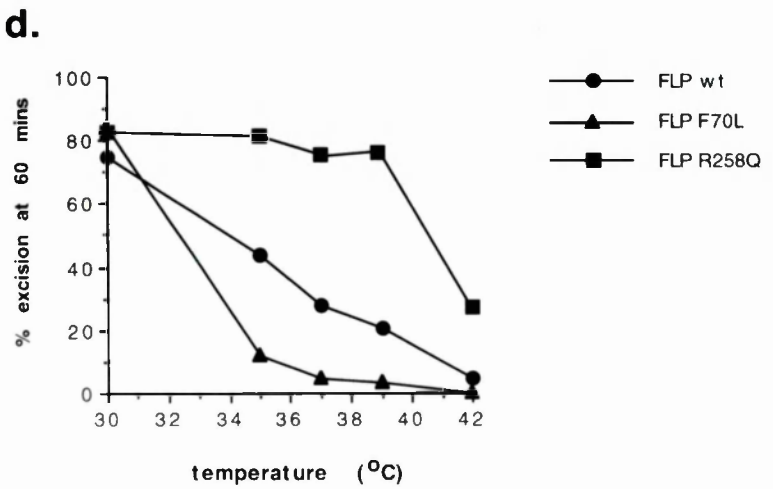
In a genetic screen for mutations that improve the binding of FLP to its substrate, Lebreton et al. (1988) defined a point mutation that increases the affinity of FLP for its DNA both in an *in vivo* assay in *Salmonella*, and using partially purified protein *in vitro*. Interestingly, the point mutation, which changes FLP amino acid 258 from Arg to Gln was shown to have increased DNA binding activity in the *Salmonella* assay at 40°C compared to 37°C. This led to the speculation that FLP R258Q may be a temperature resistant mutant. To address this possibility, I tested *in vitro* translated FLP R258Q at different temperatures, and compared its activity to that of *in vitro* translated wild type FLP (figure 5.3c, d). As with the wild type protein, the FLP R258Q protein concentration for this experiment was determined from protein titration experiments as the minimum

Figure 5.3 Effect of temperature on recombination by *in vitro* translated FLP mutants F70L and R258Q. The amount of protein used was determined as the minimum amount of protein required for maximum recombination of 0.01nM substrate in 60 minutes at 30°C. **a)** 20pmol in vitro translated FLP wt, **b)** 20pmol FLP F70L or **c)** 12pmol FLP R258Q were incubated with 0.01nM of pSVpaZ excision substrate in 80µl volume at each of the temperatures shown, and 10µl aliquots removed at indicated time points. Unrecombined (4.0kb) and recombined (2.9kb) bands are shown on the right of the figure.

d) Quantification of a) to c). % excision after 60 mins is plotted against temperature for each of the proteins. *Circles* , FLP wt; *triangles*, FLP F70L; *squares*, FLP R258Q.

Figure 5.3


a. FLP wt
b. FLP F70L
c. FLP R258Q
2 5 10 20 60
(min. incubation)
23°C
30°C
35°C
37°C
39°C
42°C
4.0kb
2.9kb


d.
% excision at 60 mins
100
80
60
40
20
0
30
32
34
36
38
40
42
temperature (°C)
FLP wt
FLP F70L
FLP R258Q

amount required for maximum recombination of 0.04nM excision substrate at 30°C in 60 minutes. In the temperature experiment, FLP R258Q recombines approximately 80% of its substrate in 60 minutes at all temperatures except 42°C. At 42°C the recombination efficiency is reduced, reaching only 25% excision after 60 minutes (figure 5.3c, d). Thus it appears that the mutant FLP R258Q is less temperature sensitive than wild type FLP in this assay. Some temperature sensitivity is however observed. The activity of R258Q is reduced at temperatures above 35°C at lower protein inputs (data not shown). Interestingly, at optimal protein input, the initial rate of recombination observed for R258Q is not faster than for wild type FLP (compare the 5 minute time point at all temperatures for FLP wt and FLP R258Q). This shows that the ability of FLP R258Q to recombine its substrate at higher temperatures than wild type FLP in this assay is not due to a faster initial recombination rate. Taken together, these results suggest that FLP R258Q is temperature sensitive, but to a lesser extent than wild type FLP.

This result is in contrast to the result of Lebreton et al., which suggests that FLP R258Q has increased DNA binding activity at 40°C. This result may be due to the expression system which was used for the study. FLP and FLP mutants were expressed from the temperature inducible λP_R promoter, whose activity is higher at 40°C than at 37°C. Thus the increased activity of FLP R258Q at the higher temperature probably results from a combination of increased expression levels and reduced sensitivity to temperature.

The effects of DNA binding on FLP R258Q have not been tested due to the lack of purified protein. For *in vitro* translated proteins, because the protein of interest represents such a small proportion of the total protein present, the requirement for non -specific competitor DNA in the bandshift assay is so high that accurate quantitative analysis is not feasible. In addition, as shown in chapter 4.2, the correct determination of DNA binding constants is dependent upon accurate measurement of protein concentration, which is not easily achieved with *in vitro* translated protein preparations.

5.3 Discussion

In this chapter I have shown that in an *in vitro* excision assay, FLP has a temperature optimum near 30°C, whereas Cre is able to recombine efficiently at temperatures of up to 42°C. Others in our lab have shown that FLP and Cre have similar optimum temperatures in *E. coli* and in a mammalian cell line (Buchholz et al., 1996). I have shown that the loss of FLP activity *in vitro* at temperatures above 30°C can be to some extent overcome by increasing the protein input, and that loss of activity occurs at the level of DNA binding. Two FLP mutants are described. The mutant FLP F70L shows increased temperature sensitivity compared to wild type FLP. The mutant FLP R258Q is temperature sensitive, but to a lesser extent than wild type FLP.

The difference in temperature optima shown for FLP and Cre is consistent with their origins: FLP comes from the yeast *Saccharomyces cerevisiae*, whose growth temperature is 30°C, whereas Cre comes from bacteriophage P1. Note that the Kw recombinase described in chapter 3, is also a yeast enzyme, and also has a temperature optimum near 30°C *in vitro* (see chapter 3, section 3.2.2, figure 3.2e).

The fact that the DNA binding affinity of FLP is reduced at temperatures above 30°C, and that the loss of recombination activity at similar temperatures can be restored by increasing protein input, is consistent with the kinetic results presented in the preceding chapter. The kinetic analysis suggested that for FLP, the step of DNA binding is rate limiting at limiting protein concentrations. This result predicts that if the DNA binding affinity is further reduced by increased temperature, then more protein will be required for a given number of recombination events.

This point raises the question of whether the mutant proteins FLP F70L and R258Q in fact have an altered DNA binding affinity. For the reasons stated in section 5.2.3 of this chapter, it is not possible to accurately measure DNA binding constants of *in vitro* translated

proteins using the bandshift assay. However the recombination characteristics of the two proteins may give some indication of the DNA binding affinities. For FLP F70L, it is unlikely that the DNA binding affinity at 30°C is different to that of wild type FLP for two reasons. Firstly the kinetics of recombination of the two proteins are identical at 30°C. Secondly, the same amount of protein for both wild type and F70L is required for optimum recombination at 30°C. Although there are difficulties in determining the amount of active protein in *in vitro* translated preparations, comparison of *in vitro* translated and purified preparations, when available, shows that for both sources of protein, the amount of protein required for optimum recombination is inversely proportional to its DNA binding affinity (see chapter 4, sections 4.3.1 to 4.3.4 for a detailed discussion of this point. Compare also figures 4.14 and 4.16).

For the mutant FLP R258Q, about two fold less protein is required for optimal recombination of a given amount of substrate. This suggests that FLP R258Q may indeed have a higher DNA binding affinity than wild type FLP, and is consistent with the results of Lebreton et al. (1988) which suggested that FLP R258Q has increased DNA binding affinity in *Salmonella*, and in a bandshift assay. To further investigate DNA binding affinities of the FLP mutants described here, *in vitro* translated proteins could be tested in the substrate saturation assay described in chapter 4 (section 4.3.3).

The different thermostabilities of Cre, FLP and the FLP mutants described here have implications for their applied use in living systems. The fact that FLP loses activity at temperatures above 35°C, may limit the choice of organisms in which it can be employed for applications in which efficient recombination is required. However, the fact that the loss of activity can be overcome *in vitro* by increasing the protein input, suggests that FLP may function efficiently in organisms with high growth temperatures if a sufficiently high level of recombinase is expressed. This suggestion appears to be borne out by the recent work of Dymecki (1997) who reported FLP mediated excision of a LacZ reporter gene in a transgenic mouse, reaching levels of 40% to

80% in a wide range of tissues. These levels compare favourably to those reported for transgenic Cre mice, and furthermore, were achieved using the FLP F70L mutant. The choice of promoter under which FLP can function efficiently in mammals may be widened by the use of the FLP R258Q mutant.

For those applications in mammalian systems in which efficiency is not required, but which instead depend on tight regulation, the thermolability of FLP may be an advantage. Examples of this type of application, which require a total absence of recombination before induction, and in which low level recombination or even a single event is sufficient to produce the desired change in phenotype, are recombinase induced cell differentiation (Rossi et al, 1996), or recombinase induced tumorigenesis (Pichel et al., 1993).

Finally, the different thermostabilities of FLP and Cre have implications for the mode of induction. In flies and plants, FLP has been regulated from heat shock promoters (Golic and Lindquist, 1989, Lyznik et al., 1995), where induction is usually carried out for several hours at temperatures of 37°C or higher. The efficiency of FLP mediated recombination in these experiments may be improved by adjustment of the induction protocol to give the best combination of expression level and stability. Alternatively, for applications in which efficiency of recombination is required, and in which heat shock promoters are used, Cre may be a more suitable choice. This suggestion appears to be borne out by the work of Siegal and Hartl (1996), who expressed Cre *in Drosophila* under the dual Hsp70/mos1 promoter, showing virtually 100% excision of a white reporter gene in somatic and germline cells.

CHAPTER 6
The effect of distance between FRTs on FLP mediated excision

6.1 Introduction

All the excision experiments presented so far in this thesis have been based on constructs in which the distance between the target sites is 1.1kb. The efficiency of excisive recombination may be affected by the distance between target sites in several ways. Firstly, one would expect that the longer the distance between target sites, the slower the rate of synapsis between them. This idea has not been directly tested for FLP or Cre, but is borne out by *in vivo* data from various systems, showing that recombination between chromosomal target sites is increasingly inefficient as the distance between them increases (Osborne et al., 1995, Ramirez-Solis et al, 1995, Golic et al., 1994, Golic and Golic, 1996).

Secondly, a minimum distance would be expected, below which recombination is physically impossible due to the constraints imposed on the synaptic complex by the length of DNA between the two sites which must be brought together. Since excisive recombination requires the formation of a circle as an excision product, the minimum possible size of such a circle may place an additional constraint on the minimum length of DNA between recombining sites. For Cre, the minimum size of circles produced by *in vitro* excision is 116 base pairs (Hoess et al., 1985). For FLP, the minimum distance has not been tested.

In this chapter, the effect of distances ranging from 74 base pairs to 15 kb on FLP mediated excision *in vitro* is examined. The results demonstrate that FLP is virtually unable to excise a substrate with 74 base pairs between its target sites (measured between the centres of the spacers), and that the optimum distance, giving the fastest initial rate of recombination, is between 400 and 700 base pairs. At a given protein to substrate ratio, recombination efficiency is greatly decreased at distances of 8kb and above, but can be restored to 100% excision by increasing the protein monomer to substrate ratio. In addition, the mathematical relationship between the rate of synapsis and the distance between FRT sites is examined.

This project is part of a collaboration with Sophie Chabanis, in which the optimum distance for recombination will also be determined using chromosomally integrated constructs in 293 cells in culture, and the relationship between distance and recombination will be compared in the *in vitro* and *in vivo* experiments.

6.2 Results

6.2.1 The effect of distance between FRT sites on excision rates *in vitro*: determination of the minimum and maximum optimum distance.

To study the effect of distance between FRT sites on recombination, the FRED11 series of excision substrates was constructed (Figure 6.1a). FRED11 has 74 bp between its FRT sites (measured between the centres of the spacers). A 15 kb yeast DNA fragment was inserted between the FRT sites of FRED11, and a series of smaller inserts, ranging from 9.9 kb to 200 bp, were generated by restriction digestion or by exonuclease III deletion. For the *in vitro* excision assay, each FRED construct was linearised, radioactively labelled by end filling, and quantified as described in "Materials and Methods". As shown in figure 6.1a, FLP mediated excision of any FRED construct gives a 6.6 kb labelled linear product, and an unlabelled circle of variable size. Thus in each panel of figure 6.1d, the recombined product is represented by the 6.6kb band. For substrates with very short distances between FRTs, the difference in migration of the substrate and product bands would not be sufficient to detect recombination. Therefore the restriction strategies shown in figures 6.1b and c were designed. For the substrates with 700 bp, 400 bp and 74 bp between FRTs (figure 6.6b), upon BglII digestion, the unrecombined substrate gives two bands, whose sum indicates the quantity of unrecombined DNA. Recombination results in excision of the DNA carrying the BglII restriction site, so that labelled linear recombined products are not digested by BglII. The 6.6kb band therefore indicates the quantity of recombined DNA. In the substrate with 200 bp between its FRTs, the BglII site is not present. For this substrate the strategy shown in

Figure 6.1 Effect of distance between FRT sites on FLP mediated excisive recombination.

a) Schematic representation of the FRED11 plasmid series of FLP excision substrates, in which directly oriented FRTs (*black arrows*) are separated by varying lengths of DNA, from 74 bp to 15 kb. Each FRED11 plasmid was linearised and radioactively labelled as described in "Materials and Methods". Radioactive label is shown as asterisks. The product of recombination of all FRED11 substrates is a linear labelled molecule of 6.6kb and an unlabelled circle of varying size. **b)** Digestion strategy used to detect recombination of FRED11 plasmids with small inserts, of 700, 400 and 74 base pairs. A single BglII site is present between the FRTs of these constructs. Upon digestion of end labelled substrates with BglII, the unrecombined substrate yields two labelled fragments, whilst the recombined linear substrate is not cut, having lost its BglII site. Recombined molecules are seen as a labelled band of 6.6 kb. **c).** Digestion strategy used to detect recombination of FRED1102 (200 bp between FRTs). This construct has no unique site between the FRTs. Instead, the EcoRV site was used. Digestion of the unrecombined substrate gives two labelled bands, of 3.8 and 2.5kb. The linear recombination product gives labelled bands of 4.1 and 2.5kb.

d),e). 3.2 nM FLP was incubated with 0.1nM of each labelled FRED11 substrate (protein monomer: 1/2 site ratio, 8:1), and aliquots were removed at each of the time points shown above the figure. λ DNA digested with HindIII was added to each experiment as a carrier, such that the total DNA concentration in each experiment was 2.7ng/μl. The distance between FRTs in each construct is shown at the side of the figure. Band sizes are shown next to each panel. For all constructs except FRED1102 (200 bp between FRTs), the recombined band is 6.6kb. For FRED1102, the recombined band is 4.1kb. f) Quantification of d) and e). The % recombination at 2 minutes (*open circles*) and at 60 minutes (*closed circles*) is plotted against distance between FRTs for each FRED11 construct.

Figure 6.1

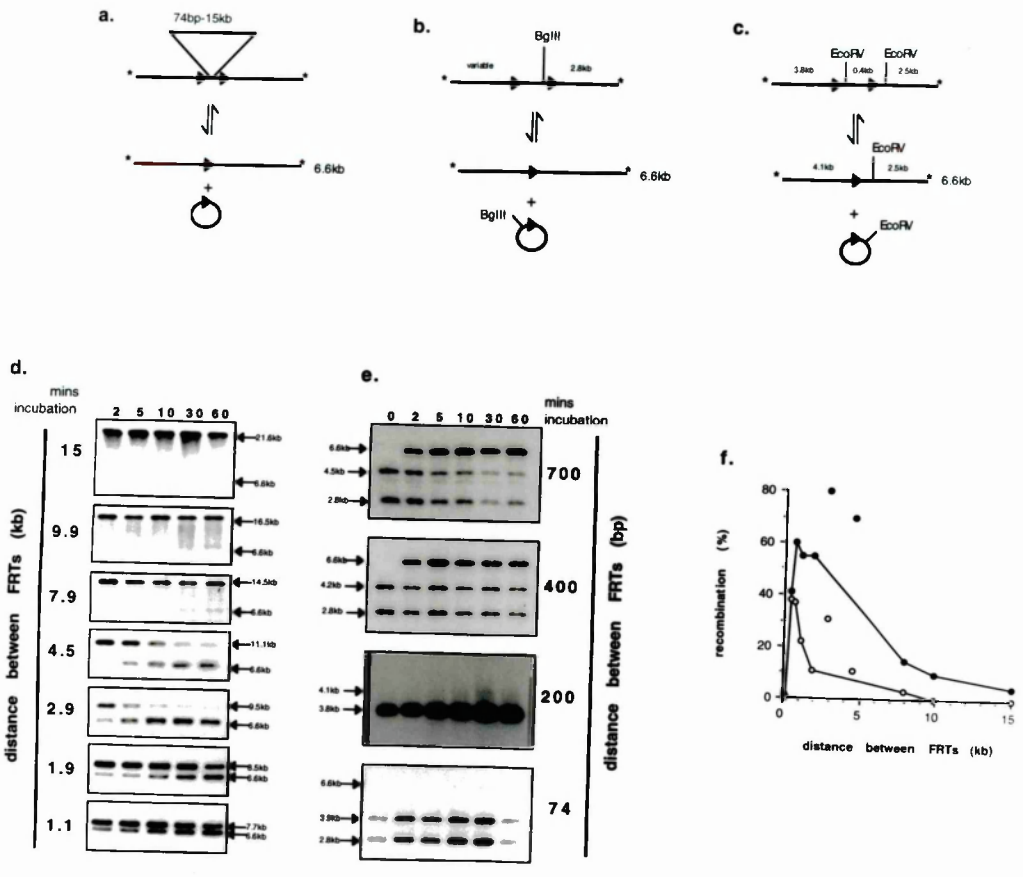

a.
74bp-15kb
6.6kb
b.
BglII
6.6kb
BglII
c.
EcoRV
EcoRV
EcoRV
6.6kb
EcoRV
d.
mins incubation
2 5 10 30 60
distance between FRTs (kb)
15
9.9
7.9
4.5
2.9
1.9
1.1
e.
0 2 5 10 30 60
mins incubation
700
400
200
74
distance between FRTs (bp)
f.
recombination (%)
distance between FRTs (kb)

figure 6.1c was designed. Upon digestion with EcoRV, the unrecombined substrate gives three fragments, two of which are labelled. Recombination results in the excision of DNA carrying one EcoRV site. The EcoRV digested linear recombination product gives two bands. The ratio of the 3.8 kb (substrate) to the 4.1 kb (product) bands gives the ratio of substrate to product.

To examine the effect of distance between FRTs on recombination rates, FLP was incubated with each FRED construct. The molar ratio of FLP monomer to FRT half site was the same in each experiment, (8:1) and the resulting difference in total DNA concentration caused by the difference in the sizes of substrate plasmids was compensated for by adding λ DNA as a carrier, such that each reaction contained 2.7ng/μl DNA (corresponding to the total DNA in 0.2nM of the largest construct, FRED11 15, whose size is 21 kb). For each FRED construct, a time course of recombination was performed (figure 6.1d, e.)

The time course of recombination for distances of 15 kb, 9.9 kb and 7.9 kb shows that recombination is inefficient at these distances under these conditions. All three substrates are less than 15% recombined after one hour. (figure 6.1f). For the very short distances, 200 base pairs and 74 base pairs, the total recombination after one hour is less than 2%. In experiments in which the protein input is 3 fold higher, recombination of the 74 base pair substrate remains below 2%, whilst the 200 base pair substrate is 70% recombined (data not shown). The effect of increasing the distance between FRTs from 200 to 400 base pairs is striking: The initial rate of recombination is increased dramatically, (38% in two minutes, figure 6.1e, f). However the total recombination does not rise above 41% after one hour. An increase in distance from 400 to 700 base pairs does not have a significant effect on the initial rate of recombination, (37% in two minutes, figure 6.1e,f) but increases the total recombination achieved after one hour to 60%. Further increases in distance, to 1.1 kb and 1.9 kb, substantially reduce the initial rate of recombination (22% and 12% in two minutes for 1.1 and 1.9 kb respectively) whilst the total recombination seen after one hour is not significantly

affected (figure 6.1f, compare also the time course data in panel d for 1.1 and 1.9 kb).

The data points for the substrates with 2.9 kb and 4.5 kb between FRTs lie above the curve suggested by the other data points. Evaluation of the total DNA in these experiments by PhosphorImager analysis of the sum of recombined and unrecombined bands showed that the DNA input was lower in these two experiments than in all other experiments. The DNA inputs in the 2.9 kb and 4.5 kb experiments are approximately 2 fold and 1.5 fold lower respectively than in other experiments. This means that the FLP monomer to half site ratio is 16:1 in the 2.9 kb experiment, and 12:1 in the 4.5 kb experiment. In chapter 4 (section 4.2.13) I showed that the initial rate and the total recombination after one hour are dependent on the protein :substrate ratio. Therefore I conclude that the higher levels of recombination seen in these two experiments are due to the greater molar excess of protein over substrate. If these points were corrected for the difference in substrate input, I expect that the initial rate of recombination for 2.9 kb would be slower than 1.9 kb, and for 4.5 kb, slower still. This expectation is confirmed by other experiments, see for instance, figure 6.3.

In summary these results show that under these experimental conditions, the fastest initial rate of recombination is seen when the distance between FRTs is 400 to 700 base pairs. Longer distances slow down the initial rate of recombination, whilst at distances of 200 base pairs and less, recombination is barely detectable. For 200 base pairs but not for 74 base pairs, recombination can be increased by increasing the protein input. Optimum total recombination is seen at distances of between 700bp and 2 kb. Above 8 kb, the total recombination is greatly reduced.

What is the reason for these results? The effects on the initial rate of recombination are probably due to the effect of distance on the rate of synapsis. At longer distances, the volume in which two bound sites must diffuse in order to find each other is increased. This would lead to a reduction in the rate of synapsis. At distances shorter than the optimum, steric constraints may also effectively reduce the rate at which two bound FRTs can meet. Other factors

which would begin to play a role as distance decreases would be the phasing of the two FLP dimers on the faces of the helix. Furthermore, the fact that FLP introduces a DNA bend of greater than 144 degrees upon binding to both halves of its target, introduces further conformational constraints (Schwartz and Sadowski, 1990). Experiments with distances between 400 and 74 base pairs, at shorter intervals, and at various protein inputs, would be necessary to further investigate this point.

Regarding the effects of distance on total recombination in one hour, the experiment at 400 base pairs is particularly interesting, as it represents a case in which the initial rate of recombination is optimal, whereas the total recombination is reduced compared to the 700 base pair experiment. It is probable that in this experiment, the equilibrium level of recombination is affected by a parameters other than the forward rate of synapsis. A possible explanation for this behaviour is that the dissociation of the synaptic complex may be accelerated by steric constraints imposed by the length of the DNA loop thus formed. Such a constraint may be reduced if the DNA loop is longer. This may explain why more recombination is seen at 700 bp than at 400 bp although the initial rate is similar. The effect of such a steric constraint on the dissociation of the synapse would be stronger with decreasing distance. Again, experiments with distances between 700 and 200 base pairs would be necessary to test this idea.

A further prediction of such a mechanism is that the dissociation rate of the synapse after recombination should be affected in a different manner than the dissociation before recombination, since the constraint imposed by DNA looping is removed, and replaced by the situation of binding one linear molecule and one small circle. At very small distances, the differences in topology of the DNA in the synapse pre and post recombination, may also affect the relative ease with which the first and second pair of strands is exchanged, and indeed the preferred direction of Holliday junction resolution. The choice of direction may thus rest on the energetic differences between a short loop and a small circle. This difference may itself vary depending on the length of DNA involved.

6.2.2 Mathematical modelling of the effects of distance

The data presented in the preceding section shows the relationship between distance and recombination efficiencies for a given set of experimental conditions. It is possible that this relationship may contain information about the mode of synapsis. To investigate this idea, I used the mathematical model described in chapter 4 (section 4.3.6).

In the mathematical model, the parameter k_3 (see Table 4.4) describes the forward rate of synapsis between two target sites linked on the same molecule. In the experiments shown in figure 6.2, this parameter was varied to simulate the effects of increasing the distance between these two targets. Three possible relationships between k_3 and distance (s) were examined:

1) (Figure 6.2a). If synapsis occurs by a linear process such as tracking along the DNA, the rate of synapsis will be inversely proportional to the distance between FRTs. If tracking is directional, then the relationship between the rate of synapsis (k_3) and the distance (s) between FRTs is given by

$$k_3 = n/s$$

2) (Figure 6.2b). If tracking is not directional, but occurs randomly in both directions along the DNA then the relationship between the rate of synapsis and the distance between FRTs is given by

$$k_3 = n/s^2$$

3) (Figure 6.2c). If synapsis occurs by a three dimensional process, such as random collision, then one might imagine that the distance between FRT sites will define the volume in which two bound FRTs must diffuse in order to find each other, and the rate of synapsis will be inversely related to this volume, which is itself directly related to the cube of the distance. To test this idea, recombination was simulated at various distances, using the relationship

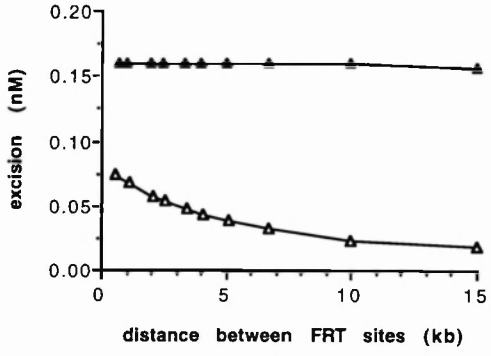
$$k_3 = n/s^3$$

Figure 6.2 Mathematical modelling of the effect of distance on recombination

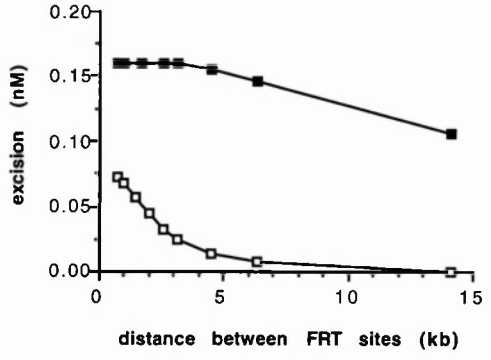
The mathematical model described in Chapter 4 (section 4.3.6) was used to simulate time course curves using the input parameters for FLP shown in table 4.5. The protein and susbtrate inputs were 4nM and 0.1 nM respectively. The parameter k3, describing the forward rate of synapsis of the fully bound substrate SM4 (see table 4.4) was varied to simulate the effects of increasing distance between FRT target sites. In each of figures a, b, and c, the nM simulated recombination at 2 minutes and at 60 minutes is plotted against the theoretical distance between FRTs. **a)** The constant k3 is assumed to be inversely proportional to the distance ($k3 = n/d$). Such a relationship would be expected if synapsis occurs by a one dimensional process (e.g., by tracking along the DNA). *Open triangles,* nM recombination at 2 minutes; *closed triangles*, nM recombination at 60 minutes. **b)** The constant k3 is assumed to be inversely proportional to the square of the distance ($k3 = n/d^2$). Such a relationship between distance and synapsis would be expected if synapsis occurs by a two dimensional process. *Open squares,* nM recombination at 2 minutes; *closed squares*, nM recombination at 60 minutes. **c)** The constant k3 is assumed to be inversely proportional to the cube of the distance ($k3 = n/d^3$). Such a relationship would be expected if synapsis occurs by a three dimensional process (e.g., by randon collision). *Open circles,* nM recombination at 2 minutes; *closed circles*, nM recombination at 60 minutes. The value of "n" was determined as that which gives the value of k3 at 1.1kb distance as used in chapter 4 (table 4.5).

Figure 6.2

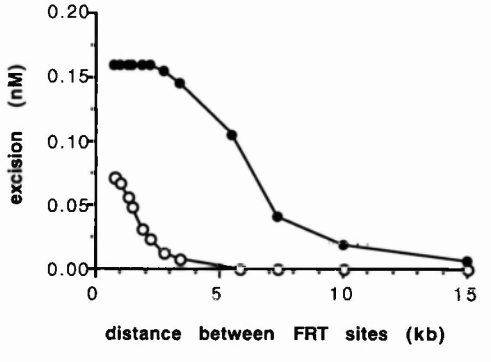
a. $k_3 = n/d$



b. $k_3 = n/d^2$



c. $k_3 = n/d^3$



However, it should be pointed out that the relationship $k_3 = n/s^3$ is an oversimplification becuase it assumes the DNA to be a fully flexible chain, and does not take into account the bending rigidity, which would affect the true flexibilty of DNA and thus the rate of synapsis. Theoretical and experimental data has enabled values for bending rigidity to be determined (persistence length, A = 450 to 500 Angstroms; approximately 150 bp. Hagerman, 1988; Frank-Kamenetskii et al., 1985). Taking persistence length into account, the relationship between synapsis time and distance (s) between two sites has been defined for random collision in supercoiled DNA substrates as inversely proportional to $s^{3/2}$ (Marko and Siggia, 1995). For linear molecules, again taking into account parameters defining DNA rigidity, the rate of encounter between two sites is also inversely proportional to $s^{3/2}$ (Rippe et al 1995). However this relationship only holds for molecules longer than 2000bp. For shorter molecules (between 100 and 1500 bp) an alternative model is generally used, in which the relationship between distance (s) and encounter rate (t) is complex, and approximates to

$$t = (n1/s^{5)} \text{ X } \exp(n2/s)$$

where n1 and n2 are constants. (Rippe et al 1995).

For simplicity, only three simple alternative models ($k_3=n/s$; $k_3=n/s^2$ and $k_3=n/s^{3)}$ were tested as shown in figure 6.2.
Comparison of the three plots in figure 6.2 with the data shown in figure 6.1f shows that the data is best described by the realtionship $k_3 = n/s^3$. Figure 6.2c shows that using this relationship, the model predicts a sharp decrease in the initial rate of recombination as the distance between FRTs increases from 0.7 to 3kb, whilst the total recombination seen at 60 minutes is relatively unaffected. The observed data (figure 6.1f) clearly follow this pattern (with the exception of the 2.9kb distance, for which recombination rates cannot be determined accurately on the basis of this data). In addition, the model predicts a severe reduction in both initial and total recombination at distances of 8kb and above (figure 6.2c), which is also observed experimentally (figure 6.1f). For the other two relationships (figures 6.2a and b) although a decrease in initial

rate is predicted with increasing distance, the decrease in total recombination is less than that observed experimentally.

Thus the results of the mathematical modelling show that the relationship between synaptic rate and distance which best describes the observed data is $k_3 = n/s^3$. This result is inconsistent with the prediction that a random collision mechanism of synapsis would be best described by the relationship $k_3 = n/s^{2/3}$ (Rippe et al 1995). However, results from topological studies of the products of FLP mediated *in vitro* recombination on supercoiled substrates, suggest that synapsis occurs exclusively by random collision (Beatty et al 1986). Thus it is likely either that additional factors are responsible for determining the observed relationship between FLP mediated recombination and distance between FRTs, or that the parameter k_3 used in the mathematical model for recombination is not sufficient to describe the synaptic rate.

6.2.3. The effect of distance is dependent on protein concentration.

The experiments in the preceding section were carried out at limiting protein concentration, at a FLP monomer to half site ratio of 8:1 at which only partial recombination takes place. How does protein concentration affect recombination over different distances? To answer this question I carried out the protein titration experiment in figure 6.3. The effect of protein concentration on recombination of three different FRED constructs, with distances of 1.1kb, 4.5kb and 9.9kb between their FRTs was compared. Each of panels a to l in figure 6.3 shows a protein titration, in which the same range of FLP concentrations is used. Each experiment was repeated at four different substrate concentrations. In this way, a wide range of protein monomer to half site ratios is represented for each substrate (from 1:1 to 160:1), but most ratios are repeated several times, enabling comparisons to be made within and between substrates. In all experiments, the total DNA concentration was made up to 2.7 ng/µl with carrier DNA.

In figure 6.3 m to p, recombination over the three different distances is compared at each substrate concentration separately. Plot m, corresponding to experiments a to c, shows that all three substrates can be recombined to near 100% by a sufficient excess of protein. Maximum recombination is reached for all three substrates at a FLP monomer to half site ratio of 80 to 1. At higher substrate inputs (figure 6.3 n, o, p) the recombination efficiencies of the three substrates differ from each other in a protein concentration dependent manner. At 32 fold molar excess of FLP over FRT half site (last points on plot n), all three substrates are more than 70% recombined. At 16:1 molar excess, (last points on plot o), the substrate with 1.1kb between FRTs (circles) is 90% recombined, whilst the 4.5kb (triangles) and 9.9kb (squares) substrates are, respectively, 70% and 58% recombined. At a FLP monomer: FRT 1/2 site ratio of 8:1, (last points, plot p), the difference in recombination efficiencies of the three substrates is accentuated, reaching 79%, 60% and 19% for the 1.1kb, 4.5kb and 9.9kb distances respectively. This effect is also evident on comparison of the three distances at different protein concentrations within each plot.

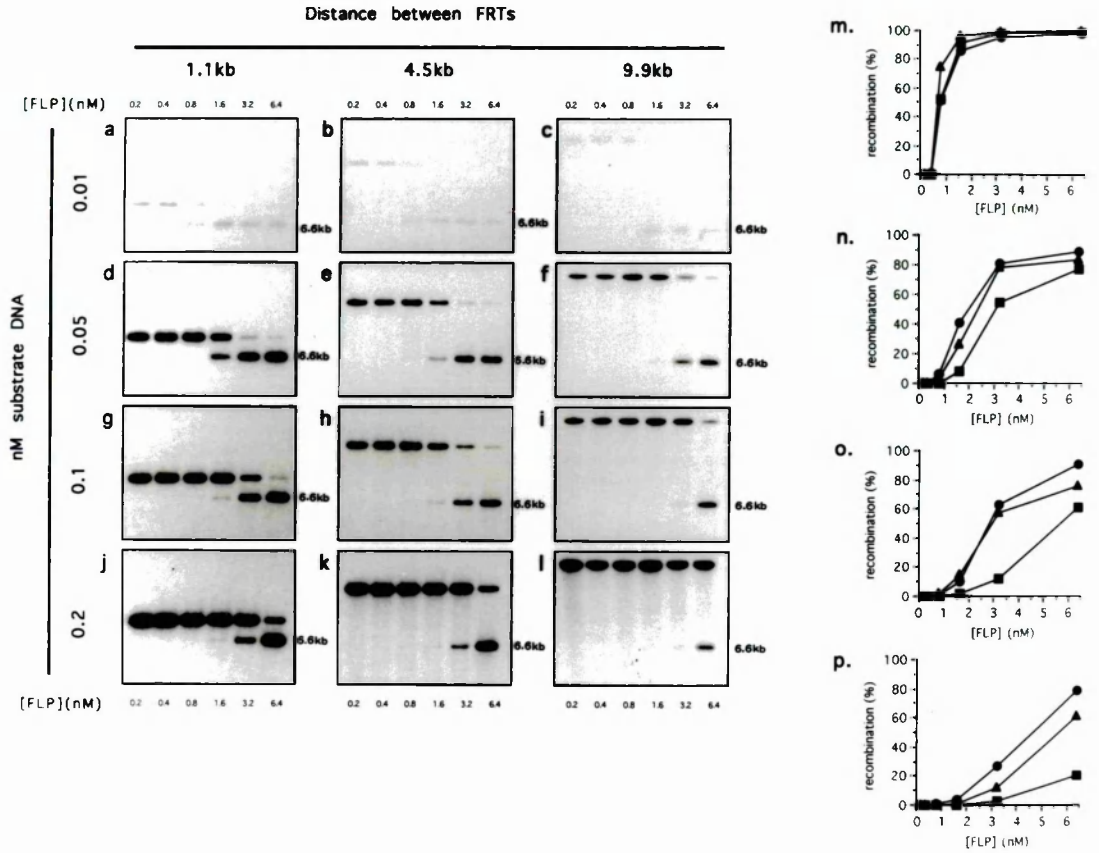
At a given protein monomer to half site ratio, the concentration of substrate DNA affects recombination efficiency over all three distances. Comparison of plots m to p at the 0.4 nM (m) 1.6 nM (n) 3.2 nM (o) and 6.4 nM (p) points , all of which represent a FLP excess of 8:1 over half site, shows that the total recombination increases with increasing substrate concentration. Note that the total DNA in each experiment was kept constant by addition of carrier DNA, suggesting that this effect is specifically related to the FLP- FRT interaction. (This effect is also discussed in Chapter 4, section 4.2.12).

In summary, these results show that the effect of distance on recombination efficiency can be overcome by sufficient protein concentration, and show that, the larger the distance between FRTs, the greater the excess of protein required for maximum recombination.

Figure 6.3 Effect of protein input and substrate input on recombination over different distances.

a) to **l)**. Various concentrations of FLP (shown above and below the figure) were incubated with three different labelled FRED11 constructs (shown above the figure) at four different substrate input concentrations (shown at the left of the figure). Reactions were incubated at 30°C for 60 minutes. In each panel, the upper band is unrecombined, and the lower band, marked 6.6kb, is the recombined product. λ DNA digested with HindIII was added as a carrier, such that the total DNA concentration in each reaction was 2.7ng/μl. **m)** to **p)**. Quantification of a) to l). % recombination is plotted against FLP concentration for each substrate input. **m)** 0.01 nM substrate. **n)** 0.05 nM substrate. **o)** 0.1 nM substrate. **p)** 0.2nM substrate. On each plot, all three FRED11 constructs are shown: *circles*, 1.1kb between FRTs. *Triangles*, 4.5kb between FRTs. *Squares*, 9.9kb between FRTs.

Figure 6.3


Distance between FRTs
1.1kb
4.5kb
9.9kb
[FLP](nM)
0.2 0.4 0.8 1.6 3.2 6.4
nM substrate DNA
0.01
0.05
0.1
0.2
a
b
c
d
e
f
g
h
i
j
k
l
6.6kb
m.
n.
o.
p.
recombination (%)
[FLP] (nM)

6.2.4. Apparent excision over very short distances occurs via intermolecular recombination

FLP is able to catalyse not only excision but also intermolecular recombination. Figure 6.4a shows how intermolecular recombination may give rise to products which are identical to the products of excision. The result of an uneven exchange between two identical excision substrates is one shorter molecule carrying a single FRT site, and one longer molecule, carrying three FRTs. As shown in figure 6.4, this longer molecule can undergo excision between the two distal FRTs, to give a product with one FRT, and a circle with two FRTs. Thus products which are indistinguishable from true excision products can be formed even in the absence of excision. For substrates in which the distance between FRT sites is too short to allow excision to take place, apparent excision products formed in this way may lead to incorrect interpretation of the results. Intermolecular recombination products themselves are not detectable in these experiments because the labelled fragments are digested to the same size as the starting substrates (see figure 6.1b,c).

To examine this phenomenon more closely, I made use of the observation, described in Chapter 4, section 4.2.12, that intermolecular recombination is enhanced by NaCl. Figure 6.4 b and c show a time course of recombination of the substrate with 1.9 kb between its FRTs, at 200mM (b) and 150mM (c) NaCl. Intermolecular recombination products resulting from uneven exchange between two substrate molecules, are longer than the starting substrate by increments of 1.9kb. Products of intermolecular recombination are detectable in the experiment at 200mM NaCl (figure 6.4b) but not at 150mM (figure 6.4c). The rate and extent of excision is relatively unaffected by the NaCl concentration (figure 6.4d). Figures 6.4e and f show a time course of recombination of the 74bp substrate at two NaCl concentrations. 5 fold more excision product is seen at 200mM NaCl than at 150mM, suggesting that at 200mM NaCl this product arises mostly from intermolecular recombination as described above. It is possible that this result merely indicates an effect of NaCl on the rigidity of DNA, but I disfavour this explanation because the effect of these two salt concentrations on excision of the 400

Figure 6.4 Intermolecular recombination: an alternative pathway giving apparent excision over very small distances.

a) Schematic representation of possible recombination reactions on excision substrate. 1). Intermolecular recombination can result in uneven exchange between two molecules, giving a product which is longer than the original substrate. The increase in length is equal to the distance between FRTs in the starting substrate. 2) Excision can take place between any pair of FRTs. The figure shows excision between the distal FRTs, giving a linear product with one FRT, and a circle with two FRTs. **b), c), d).** NaCl affects the frequency of intermolecular recombination, but has a minimal effect on excision. 2.4 nM FLP was incubated with 0.1 nM FRED111.9 (1.9kb between FRTs), in a buffer containing 200mM NaCl (b) 0r 150mM NaCl (c). Aliquots were taken at the times indicated on the figure. On the right of the figure, the starting substrate is indicated (8.5kb). The product of excision of this substrate is 6.6kb. The products of intermolecular uneven exchange are visible as bands of 10.4 and 12.3 kb in panel b. d) Quantification of b and c. The % excision is plotted against time. *Closed circles*, 200mM NaCl. *Open circles,* 150mM NaCl. **e), f).** 3.2nM FLP was incubated with 0.1nM FRED110074 (74bp between FRTs) in a buffer containing 200mM NaCl (e) or 150mM NaCl (f). (panel f is a long exposure of data in fig 6.1e, 74bp). Aliquots were removed at the time intervals indicated above the figure, and digested with BglII. In each panel, unrecombined bands (3.9 and 2.8 kb) and recombined bands (6.6kb) are indicated at the right of the figure. **g)** Quantification of e and f. The % recombination is plotted against time for each experiment. *Closed triangles*, 200mM NaCl. *Open triangles*, 150mM NaCl.

base pair substrate, in which DNA rigidity would also be expected to affect recombination efficiency is minimal (data not shown). All other experiments described in this chapter were carried out at 150mM NaCl. The minimal amount of apparent excision observed in figure 6.4f may reflect a low level of true excision events, but is more likely to represent the level of intermolecular recombination occurring. To determine whether excision occurs over the shortest distance of 74 bp, it would be necessary to determine the size of the smallest circle formed, which is not possible on the basis of this assay, in which the circle is not labelled.

6.3 Discussion

In this chapter I have characterised the effect of distance between FRT sites on FLP mediated excision *in vitro,* showing that recombination is inhibited at a distance of 74 bp, that at limiting protein concentration, the optimum distance for initial recombination is between 400 and 700 base pairs, and that recombination is greatly reduced at distances above 8 kb under these conditions. I have investigated parameters which affect recombination at suboptimal distances, showing that the inhibition of recombination at 200 bp, and at 4.5 kb and 9.9 kb can be overcome by increasing the protein input, and that apparent recombination over very short distances can occur via the alternative pathway of intermolecular recombination.

No systematic evaluation of the effects of distance on recombination by FLP has been published. The minimum possible distance for FLP recombination (between 74 and 200 base pairs) is in agreement with that shown for Cre (116 base pairs, Hoess et al., 1985). Further experiments with distances between 74 and 200 base pairs will be necessary to determine the true minimum distance. Once this minimum distance is defined, it may be possible to examine the effect of varying the position of one FLP dimer with respect to the other around the faces of the helix by designing constructs with differences of less than 10 base pairs.

The inhibition of recombination seen at suboptimal distances can be overcome by increasing the protein monomer to substrate ratio. This has important implications for the interpretation of the results presented here, because variations in the substrate concentration also have the effect of changing the protein monomer to substrate ratio. When comparing different substrates to each other, accurate quantification of substrates is vital to the reliability of the results. Care was taken to control each step of the substrate labelling and recovery procedures. However some variation remains, which can be determined by quantification of total DNA in a given experiment after recombination. Variations in substrate concentration of up to about 10% cannot however be avoided, thus results in which slight variations in recombination efficiencies occur with different substrates should be interpreted with caution. The distance dependent variation presented in figure 6.2 is considered to be significant for several reasons. Firstly, the data from different substrates (with the exception of 2.9 and 4.5 kb) are consistent with each other. A random variation in substrate concentration would be not be expected to produce a smooth decline in recombination efficiencies over increasing distance. Secondly, quantification of the total DNA in each experiment showed variations of around 10% between all substrates, with the exception of the 2.9 and 4.5 kb substrates. From this analysis I conclude that the curves shown in figure 6.1 are sufficiently correct to draw conclusions as to their general shape, but that slight variations in substrate concentration may lead to inaccuracies in the determination of the actual amount of recombination. Results from substrate titration experiments (see for example figure 4.15), suggest that a 10% variation in substrate concentration can give about 5% variation in initial and final recombination levels.

If this variation is taken into account, the curves shown in figure 6.1 are nevertheless sufficiently clearly defined for the mathematical relationship between synapsis and distance to be discernable. The fact that the mechanism of synapsis may be mathematically related to response to increasing distance (figure 6.2), raises the interesting possibility that the mechanism of Cre synapsis could be probed by the approach taken here.

Topological studies *in vitro* suggest that Cre carries out synapsis, sometimes by random collision, but mostly by an ordered mechanism, such as tracking (Abremski and Hoess, 1985). Conversely, the same authors report that in experiments in which Cre is given a substrate with four targets, proximal pairs of targets are not recombined in preference to distal pairs, suggesting that tracking is not preferred to random collision. Further complicating the picture is the study of Adams et al., (1992) in *E.coli*, also using topological assays, which suggests that *in vivo*, synapsis occurs almost exclusively by an ordered mechanism (although the same authors show that intermolecular recombination can also take place under the same conditions, showing that random collision can occur).

With the exception of the experiment in which Cre is presented with several target sites (Abremski and Hoess, 1985), all of the above results rely for their interpretation on the assumption that the topological state of reaction products faithfully reports the mechanism of synapsis. Linked reaction products are assumed to result from a random collision mechanism, whereas unlinked reaction products are considered to be the signature of an "ordered" synapsis, in which the tracking of a bound recombinase dimer along the DNA until it meets its partner, is thought to exclude supercoils (see Chapter 1, section 1.3.2 for a detailed discussion of this concept). The results with Cre so far reported do not exclude the possibility that Cre carries out synapsis by random collision, but subsequently decatenates all or some of its products by a topoisomerase like activity. An alternative test for synapsis based on the determination of the mathematical relationship between distance between LoxP sites and Cre mediated recombination *in vitro,* may help to throw light on the mechanism of synapsis.

The fact that *in vitro*, an increase in protein concentration can overcome the reduction in recombination rates caused by increasing the distance between FRTs, introduces an element of flexibility into the concept of "the maximum optimum distance for recombination". Over the range of distances tested here, the maximum distance at which recombination can occur efficiently depends on whether the protein concentration is limiting or not. This effect can be imagined

as a balance between the population of fully bound excision substrates, and the time taken for such a complex to undergo synapsis. The rate of synapsis will be reduced with increasing distance, but the population of occupied substrates will be increased with increasing protein concentration. The fact that a higher excess of protein is required for efficient recombination, the larger the distance between FRTs, (figure 6.3) is consistent with this reasoning.

Is there an upper limit to the distance at which maximal recombination can be achieved by increasing the protein input? Such and upper limit may be reached when the rate of synapsis between two sites on the same molecule approaches the rate of synapsis between sites on separate molecules, and thus once the reaction reaches equilibrium, the rate of excision may be expected to be equal to the rate of reinsertion of the excised circle, so that 100% excision is never reached. In this context it is interesting to note that in the protein titration experiment presented in figure 6.3, maximal recombination of all constructs was reached only at the lowest substrate concentration. Thus inhibition of intermolecular recombination due to substrate dilution may be a factor which contributes to this result.

Although the *in vitro* results presented here may not accurately predict the behaviour of FLP *in vivo*, certain general points emerge, which may have implications for applied site specific recombination. Firstly, the interrelationship between protein concentration, recombination efficiency and distance which I have demonstrated *in vitro*, should also operate *in vivo*. Thus, if excision is to take place over long distances, then higher levels of FLP expression should increase efficiency. Secondly, the suggestion that there may be a distance at which the rate of synapsis for excision is equal to that for reintegration, has particular relevance for site specific chromosomal integration, where one of the main impediments to the stability of integrated DNA is its reexcision. If the excision reaction could be rendered sufficiently inefficient by the distance between the target sites, then the stability of integrated constructs may be improved.

This project is part of a larger project, in which the ultimate aim is to determine the effect of distance on recombination rates on chromosomally integrated targets in mammalian cell culture, where chromatin structure may be expected to alter the effective distance between target sites, and indeed the mechanism of FLP synapsis may be different (Adams et al, 1992). We hope that such an analysis will enable conclusions to be drawn as to the maximum and minimum distances which allow efficient FLP mediated recombination *in vivo*, thus enabling the optimal design of genomic manipulation experiments. It will be interesting to see whether the effect of protein concentration which I have demonstrated *in vitro* also operates *in vivo,* i.e., whether recombination efficiencies over longer distances are improved by increasing the expression levels of FLP protein, since this would have direct relevance to experimental design.

7. SUMMARY AND FINAL CONCLUSIONS

The aim of this thesis, as stated in the first chapter, was to investigate and compare the biochemical properties of site specific recombinases, with a view to optimising their use in genomic manipulation strategies.

To this end, I have characterised a novel yeast family recombinase, the Kw recombinase from the yeast *Kluyveromyces waltii*, and have shown that it shares the features of FLP and Cre which make them particularly useful for genomic manipulation applications, namely: the Kw protein is alone sufficient to carry out both excision and intermolecular recombination, on a short target site consisting of an 18 base pair inverted repeat with a 7 base pair spacer. I have demonstrated that recombination is conservative, and have determined the optimum conditions for *in vitro* recombination, including the demonstration that Kw recombines its target site optimally at 30°C, and is inhibited at higher temperatures. Preliminary experiments in 293 cells demonstrate that Kw is able to recombine its substrate in mammalian cell culture. Both *in vitro* and *in vivo*, Kw and FLP have distinct target site specificities. Taken together these results show that Kw possesses all the features necessary for it to be useful in genomic manipulation applications.

Kinetic analysis of FLP and Cre shows that Cre has a higher DNA binding affinity than FLP, and predicts that the rate of synapsis and the stability of the synaptic complex is higher for Cre than for FLP. These results have implications for the optimal use of recombinases for particular *in vivo* applications.

Using the kinetic data, I have established mathematical models describing both DNA binding and the full recombination reaction. Such models not only enable accurate values of rate constants governing the processes of DNA binding and some of those governing recombination to be determined, but also enable concrete predictions to be made regarding the behaviour of recombination intermediates, such as the synaptic complex. Such predictions can be tested experimentally. If the parameters used to model FLP and

Cre recombination are validated by experimental testing, then the model can further be used to predict the effect of mutations which change a given parameter, such as for example, the rate of synapsis. Such predictions may be useful in the search for mutant recombinases with altered properties.

I have shown that FLP and Cre have different optimum temperatures for recombination *in vitro.* FLP loses activity at temperatures above 35°C, whereas Cre is active at all temperatures tested, up to 42°C. This result has important implications for the applied use of FLP and Cre in vivo, because the growth temperature of the host organism imposes a condition on the recombinase. The optimum temperature for Kw recombination in vitro is also 30°C. In addition, kinetic analysis suggests that Kw has similar DNA binding affinity and general recombination kinetics to FLP. These results suggest that in experiments in which two recombinases are used simultaneously to recombine distinct targets, FLP and Kw may both function with similar efficiencies in the same environment.

Finally, I have studied the effect of distance between FRT sites on FLP mediated excisive recombination *in vitro.* My results show that FLP is virtually unable to excise at a distance of 74 bp, measured between the centres of the FRT spacers. At a given protein: substrate ratio, the optimum distance for the initial recombination rate is 400 to 700 base pairs, and recombination is severely reduced at distances of 7.9 kb and above. The range of distances over which efficient recombination can occur is increased by increasing the protein input.

The effect of FLP protein concentration is a recurring theme in this thesis. For example, not only the effects of distance, but also the effects of temperature on FLP activity can be to some extent alleviated by increasing the protein input. The main consequence of the low DNA binding affinity measured for FLP is to introduce a requirement for a high concentration of protein, much of which is probably not actually bound to an FRT target at any one time, but whose presence may increase the probability of encounters between FLP and its target, thereby increasing the likelihood that all four

half sites will be simultaneously filled, and that synapsis can take place.

The results of the *in vitro* and mathematical analyses presented here show that the FLP, Kw and Cre recombinases have inherently different properties, some of which will be important in determining their optimal use in *in vivo* experiments. Which of these properties are relevant to their behaviour in a cellular environment? On the basis of this study, *in vivo* experiments can be designed to answer this question.

References

Abremski, K., Frommer, B., and Hoess, R. H. (1986b). Linking-number changes in the DNA substrate during Cre-mediated loxP site-specific recombination. *J Mol Biol 192*, 17-26.

Abremski, K., and Hoess, R. (1984). Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J Biol Chem 259*, 1509-14.

Abremski, K., and Hoess, R. (1985). Phage P1 Cre-loxP site-specific recombination. Effects of DNA supercoiling on catenation and knotting of recombinant products. *J Mol Biol 184*, 211-20.

Abremski, K., Hoess, R., and Sternberg, N. (1983). Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. *Cell 32*, 1301-11.

Abremski, K., Wierzbicki, A., Frommer, B., and Hoess, R. H. (1986a). Bacteriophage P1 Cre-loxP site-specific recombination. Site-specific DNA topoisomerase activity of the Cre recombination protein. *J Biol Chem 261*, 391-6.

Abremski, K. E., and Hoess, R. H. (1992). Evidence for a second conserved arginine residue in the integrase family of recombination proteins. *Protein Eng 5*, 87-91.

Adams, D. E., Bliska, J. B., and Cozzarelli, N. R. (1992). Cre-lox recombination in Escherichia coli cells. Mechanistic differences from the in vitro reaction. *J Mol Biol 226*, 661-73.

Albert, H., Dale, E. C., Lee, E., and Ow, D. W. (1995). Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. *Plant J 7*, 649-59.

Amin, A. A., Beatty, L. G., and Sadowski, P. D. (1990). Synaptic intermediates promoted by the FLP recombinase. *J Mol Biol 214*, 55-72.

Amin, A. A., and Sadowski, P. D. (1989). Synthesis of an enzymatically active FLP recombinase in vitro: search for a DNA-binding domain. *Mol Cell Biol 9*, 1987-95.

Andrews, B. J., Beatty, L. G., and Sadowski, P. D. (1987). Isolation of intermediates in the binding of the FLP recombinase of the yeast plasmid 2-micron circle to its target sequence. *J Mol Biol 193*, 345-58.

Andrews, B. J., McLeod, M., Broach, J., and Sadowski, P. D. (1986). Interaction of the FLP recombinase of the Saccharomyces cerevisiae 2 micron plasmid with mutated target sequences. *Mol Cell Biol 6*, 2482-9.

Araki, H., Nakanishi, N., Evans, B. R., Matsuzaki, H., Jayaram, M., and Oshima, Y. (1992). Site-specific recombinase, R, encoded by yeast plasmid pSR1. *J Mol Biol 225*, 25-37.

Argos, P., Landy, A., Abremski, K., Egan, J. B., Haggard Ljungquist, E., Hoess, R. H., Kahn, M. L., Kalionis, B., Narayana, S. V., Pierson, L. S. d., and et al. (1986). The integrase family of site-specific recombinases: regional similarities and global diversity. *Embo J 5*, 433-40.

Babineau, D., Vetter, D., Andrews, B. J., Gronostajski, R. M., Proteau, G. A., Beatty, L. G., and Sadowski, P. D. (1985). The FLP protein of the 2-micron plasmid of yeast. Purification of the protein from Escherichia coli cells expressing the cloned FLP gene. *J Biol Chem 260*, 12313-9.

Barinaga, M. (1994). Knockout mice: round two [news; comment] [published erratum appears in Science 1994 Aug 12;265(5174):855]. *Science 265*, 26-8.

Baubonis, W., and Sauer, B. (1993). Genomic targeting with purified Cre recombinase. *Nucleic Acids Res 21*, 2025-9.

Beatty, L. G., Babineau Clary, D., Hogrefe, C., and Sadowski, P. D. (1986). FLP site-specific recombinase of yeast 2-micron plasmid. Topological features of the reaction. *J Mol Biol 188*, 529-44.

Beatty, L. G., and Sadowski, P. D. (1988). The mechanism of loading of the FLP recombinase onto its DNA target sequence. *J Mol Biol 204*, 283-94.

Benjamin, H. A. a. C., N.R. (1986). DNA- directed synapsis in recombination : slithering and random collision of sites. *Proc. R.A. Welch Found. Conf. Chem. Res. 29*, 107-126.

Benjamin, H. W., Matzuk, M.M., Krasnow, M.A. and Cozzarelli, N.R. (1985). Recombination site selection by Tn3 resolvase: topological tests of a tracking mechanism. *Cell 40*, 147-158.

Bernstein, R. M., Schluter, S. F., Bernstein, H., and Marchalonis, J. J. (1996). Primordial emergence of the recombination activating gene 1 (RAG1): sequence of the complete shark gene indicates homology to microbial integrases. *Proc Natl Acad Sci U S A 93*, 9454-9.

Blakely, G., May, G., McCulloch, R., Arciszewska, L. K., Burke, M., Lovett, S. T., and Sherratt, D. J. (1993). Two related recombinases are required for site-specific recombination at dif and cer in E. coli K12. *Cell 75*, 351-61.

Blakely, G., and Sherratt, D. (1996). Determinants of selectivity in Xer site-specific recombination. *Genes Dev 10*, 762-73.

Blakely, G. W., and Sherratt, D. J. (1994). Interactions of the site-specific recombinases XerC and XerD with the recombination site dif. *Nucleic Acids Res 22*, 5613-20.

Boocock, M. R., Zhu, X., and Grindley, N. D. (1995). Catalytic residues of gamma delta resolvase act in cis. *Embo J 14*, 5129-40.

Bruckner, R. C., and Cox, M. M. (1986). Specific contacts between the FLP protein of the yeast 2-micron plasmid and its recombination site. *J Biol Chem 261*, 11798-807.

Buchholz, F., Ringrose, L., Angrand, P. O., Rossi, F., and Stewart, A. F. (1996). Different thermostabilities of FLP and Cre recombinases: implications for applied site-specific recombination. *Nucleic Acids Res 24*, 4256-62.

Calero, S., Fernandez de Henestrosa, A. R., and Barbe, J. (1994). Molecular cloning, sequence and regulation of expression of the recA gene of the phototrophic bacterium Rhodobacter sphaeroides. *Mol Gen Genet 242*, 116-20.

Campbell, A. M. (1962). Episomes. *Adv. Genet. 11*, 101 - 145.

Chen, J. W., Evans, B. R., Yang, S. H., Araki, H., Oshima, Y., and Jayaram, M. (1992b). Functional analysis of box I mutations in yeast site-specific recombinases Flp and R: pairwise complementation with recombinase variants lacking the active-site tyrosine. *Mol Cell Biol 12*, 3757-65.

Chen, J. W., Evans, B. R., Yang, S. H., Teplow, D. B., and Jayaram, M. (1991). Domain of a yeast site-specific recombinase (Flp) that recognizes its target site. *Proc Natl Acad Sci U S A 88*, 5944-8.

Chen, J. W., Lee, J., and Jayaram, M. (1992c). DNA cleavage in trans by the active site tyrosine during Flp recombination: switching protein partners before exchanging strands. *Cell 69*, 647-58.

Chen, J. W., Yang, S. H., and Jayaram, M. (1993). Tests for the fractional active-site model in Flp site-specific recombination. Assembly of a functional recombination complex in half-site and full-site strand transfer. *J Biol Chem 268*, 14417-25.

Chen, X. J., Cong, Y. S., Wesolowski Louvel, M., Li, Y. Y., and Fukuhara, H. (1992a). Characterization of a circular plasmid from the yeast Kluyveromyces waltii. *J Gen Microbiol 138*, 337-45.

Chen, X. J., Saliola, M., Falcone, C., Bianchi, M. M., and Fukuhara, H. (1986). Sequence organization of the circular plasmid pKD1 from the yeast Kluyveromyces drosophilarum. *Nucleic Acids Res 14*, 4471-81.

Cowart, M., Benkovic, S. J., and Nash, H. A. (1991). Behavior of a cross-linked attachment site: testing the role of branch migration in site-specific recombination. *J Mol Biol 220*, 621-9.

Craig, N. L. (1988). The mechanism of conservative site-specific recombination. *Annu Rev Genet 22*, 77-105.

Craigie, R., and Mizuuchi, K. (1986). Role of DNA topology in Mu transposition: mechanism of sensing the relative orientation of two DNA segments. *Cell 45*, 793-800.

Dale, E. C., and Ow, D. W. (1990). Intra- and intermolecular site-specific recombination in plant cells mediated by bacteriophage P1 recombinase. *Gene 91*, 79-85.

Dixon, J. E., and Sadowski, P. D. (1994). Resolution of immobile chi structures by the FLP recombinase of 2 microns plasmid. *J Mol Biol 243*, 199-207.

Dixon, J. E., and Sadowski, P. D. (1993). Resolution of synthetic chi structures by the FLP site-specific recombinase. *J Mol Biol 234*, 522-33.

Dorgai, L., Yagil, E., and Weisberg, R. A. (1995). Identifying determinants of recombination specificity: construction and characterization of mutant bacteriophage integrases. *J Mol Biol 252*, 178-88.

Dunderdale, H. J., and West, S. C. (1994). Recombination genes and proteins. *Curr Opin Genet Dev 4*, 221-8.

Dymecki, S. M. (1996). A modular set of Flp, FRT and lacZ fusion vectors for manipulating genes by site-specific recombination. *Gene 171*, 197-201.

Dymecki, S. M. (1996). Flp recombinase promotes site-specific DNA recombination in embryonic stem cells and transgenic mice. *Proc Natl Acad Sci U S A 93*, 6191-6.

Ebenhöh, W. (1975). Mathematik für Biologen und Mediziner Heidelberg: Quelle und Meyer.

Evans, B. R., Chen, J. W., Parsons, R. L., Bauer, T. K., Teplow, D. B., and Jayaram, M. (1990). Identification of the active site tyrosine of Flp recombinase. Possible relevance of its location to the mechanism of recombination [published erratum appears in J Biol Chem 1991 Apr 15;266(11):7312]. *J Biol Chem 265*, 18504-10.

Feil, R., Brocard, J., Mascrez, B., LeMeur, M., Metzger, D., and Chambon, P. (1996). Ligand-activated site-specific recombination in mice. *Proc Natl Acad Sci U S A 93*, 10887-90.

Friesen, H., and Sadowski, P. D. (1992). Mutagenesis of a conserved region of the gene encoding the FLP recombinase of Saccharomyces cerevisiae. A role for arginine 191 in binding and ligation. *J Mol Biol 225*, 313-26.

Gates, C. A., and Cox, M. M. (1988). FLP recombinase is an enzyme. *Proc Natl Acad Sci U S A 85*, 4628-32.

Gellert, M. (1992). Molecular analysis of V(D)J recombination. *Annu. Rev. Genet 22*, 425 - 446.

Gellert, M. (1992). V(D)J recombination gets a break. *Trends. Genet. 8*, 408 - 412.

Gibson, T. J., Hyvonen, M., Musacchio, A., Saraste, M., and Birney, E. (1994). PH domain: the first anniversary. *Trends Biochem Sci 19*, 349-53.

Golic, K. G. (1991). Site-specific recombination between homologous chromosomes in Drosophila. *Science 252*, 958-61.

Golic, K. G., and Golic, M. M. (1996). Engineering the Drosophila genome: chromosome rearrangements by design. *Genetics 144*, 1693-711.

Golic, K. G., and Lindquist, S. (1989). The FLP recombinase of yeast catalyzes site-specific recombination in the Drosophila genome. *Cell 59*, 499-509.

Govind, N. S. a. J., M. (1987). Rapid localisation and characterisation of mutations within the 2 micron circle site specific recombinase: a general strategy for analysis of protein function. *Gene 51*, 31-41.

Gu, H., Marth, J. D., Orban, P. C., Mossmann, H., and Rajewsky, K. (1994). Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting [see comments]. *Science 265*, 103-6.

Halligan, B. D., Teng, M., Guilliams, T. G., Nauert, J. B., and Halligan, N. L. (1995). Cloning of the murine cDNA encoding VDJP, a protein homologous to the large subunit of replication factor C and bacterial DNA ligases. *Gene 161*, 217-22.

Han, Y. W., Gumport, R. I., and Gardner, J. F. (1993). Complementation of bacteriophage lambda integrase mutants: evidence for an intersubunit active site. *Embo J 12*, 4577-84.

Hernalsteens, J. P., Van Vliet, F., De Beuckeleer, M., Depicker, A., Engler, G., Lemmers, M., Holsters, M., Van Montagu, M., and Schell, J. (1992). The Agrobacterium tumefaciens Ti plasmid as a host vector system for introducing foreign DNA in plant cells. 1980 [classical article]. *Biotechnology 24*, 374-6.

Hickman, A. B., Waninger, S., Scocca, J.J. and Dyda, F. (1997). Molecular organisation of site specific recombination: The catalytic domain of bacteriophage HP1 integrase at 2.7 angstrom resolution. *Cell 89*, 227 - 238.

Hoess, R., Abremski, K., Irwin, S., Kendall, M., and Mack, A. (1990). DNA specificity of the Cre recombinase resides in the 25 kDa carboxyl domain of the protein. *J Mol Biol 216*, 873-82.

Hoess, R., Abremski, K., and Sternberg, N. (1984). The nature of the interaction of the P1 recombinase Cre with the recombining site loxP. *Cold Spring Harb Symp Quant Biol 49*, 761-8.

Hoess, R., Wierzbicki, A., and Abremski, K. (1985). Formation of small circular DNA molecules via an in vitro site-specific recombination system. *Gene 40*, 325-9.

Hoess, R., Wierzbicki, A., and Abremski, K. (1987). Isolation and characterization of intermediates in site-specific recombination. *Proc Natl Acad Sci U S A 84*, 6840-4.

Hoess, R. H., and Abremski, K. (1984). Interaction of the bacteriophage P1 recombinase Cre with the recombining site loxP. *Proc Natl Acad Sci U S A 81*, 1026-9.

Hoess, R. H., and Abremski, K. (1985). Mechanism of strand cleavage and exchange in the Cre-lox site-specific recombination system. *J Mol Biol 181*, 351-62.

Hoess, R. H., Wierzbicki, A., and Abremski, K. (1986). The role of the loxP spacer region in P1 site-specific recombination. *Nucleic Acids Res 14*, 2287-300.

Horton, R. M., Hunt, H.D., Ho, S. N., Pullen., J.K. and Pease, L.R. (1989). Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene 77*, 61-68.

Jayaram, M. (1994). Phosphoryl transfer in Flp recombination: a template for strand transfer mechanisms. *Trends Biochem Sci 19*, 78-82.

Jayaram, M. (1985). Two-micrometer circle site-specific recombination: the minimal substrate and the possible role of flanking sequences. *Proc Natl Acad Sci U S A 82*, 5875-9.

Jayaram, M., Crain, K. L., Parsons, R. L., and Harshey, R. M. (1988). Holliday junctions in FLP recombination: resolution by step-arrest mutants of FLP protein. *Proc Natl Acad Sci U S A 85*, 7902-6.

Johnson, R. C. (1991). Mechanism of site-specific DNA inversion in bacteria. *Curr Opin Genet Dev 1*, 404-11.

Kanaar, R., Klippel, A., Shekhtman, E., Dungan, J. M., Kahmann, R., and Cozzarelli, N. R. (1990). Processive recombination by the phage Mu

Gin system: implications for the mechanisms of DNA strand exchange, DNA site alignment, and enhancer action. *Cell 62*, 353-66.

Karess, R. E., and Rubin, G. M. (1984). Analysis of P transposable element functions in Drosophila. *Cell 38*, 135-46.

Kellendonk, C., Tronche, F., Monaghan, A. P., Angrand, P. O., Stewart, F., and Schutz, G. (1996). Regulation of Cre recombinase activity by the synthetic steroid RU 486. *Nucleic Acids Res 24*, 1404-11.

Kilby, N. J., Davies, G. J., and Snaith, M. R. (1995). FLP recombinase in transgenic plants: constitutive activity in stably transformed tobacco and generation of marked cell clones in Arabidopsis. *Plant J 8*, 637-52.

Kilby, N. J., Snaith, M. R., and Murray, J. A. (1993). Site-specific recombinases: tools for genome engineering. *Trends Genet 9*, 413-21.

Kimball, A. S., Lee, J., Jayaram, M., and Tullius, T. D. (1993). Sequence-specific cleavage of DNA via nucleophilic attack of hydrogen peroxide, assisted by Flp recombinase. *Biochemistry 32*, 4698-701.

Kitts, P. A., and Nash, H. A. (1987). Homology-dependent interactions in phage lambda site-specific recombination. *Nature 329*, 346-8.

Klippel, A., Cloppenborg, K., and Kahmann, R. (1988). Isolation and characterization of unusual gin mutants. *Embo J 7*, 3983-9.

Konsolaki, M., Sanicola, M., Kozlova, T., Liu, V., Arca, B., Savakis, C., Gelbart, W. M., and Kafatos, F. C. (1992). FLP-mediated intermolecular recombination in the cytoplasm of Drosophila embryos. *New Biol 4*, 551-7.

Kosman, D., and Small, S. (1997). Concentration-dependent patterning by an ectopic expression domain of the Drosophila gap gene knirps. *Development 124*, 1343-54.

Kowalczykowski, S. C. (1994). In vitro reconstitution of homologous recombination reactions. *Experientia 50*, 204-15.

Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res 12*, 857-72.

Kuhn, R., Schwenk, F., Aguet, M., and Rajewsky, K. (1995). Inducible gene targeting in mice. *Science 269*, 1427-9.

Kwon, H. J., Tirumalai, R., Landy, A., and Ellenberger, T. (1997). Flexibility in DNA recombination: structure of the lambda integrase catalytic core [see comments]. *Science 276*, 126-31.

Landy, A. (1993). Mechanistic and structural complexity in the site-specific recombination pathways of Int and FLP. *Curr Opin Genet Dev 3*, 699-707.

Lasko, M., Pichel, J. G., Gorman, J. R., Sauer, B., Okamoto, Y., Lee, E., Alt, F. W., and Westphal, H. (1996). Efficient in vivo manipulation of mouse genomic sequences at the zygote stage. *Proc Natl Acad Sci U S A 93*, 5860-5.

Lebreton, B., Prasad, P. V., Jayaram, M., and Youderian, P. (1988). Mutations that improve the binding of yeast FLP recombinase to its substrate. *Genetics 118*, 393-400.

Lee, J., and Jayaram, M. (1995b). Functional roles of individual recombinase monomers in strand breakage and strand union during site-specific DNA recombination. *J Biol Chem 270*, 23203-11.

Lee, J., and Jayaram, M. (1993). Mechanism of site-specific recombination. Logic of assembling recombinase catalytic site from fractional active sites. *J Biol Chem 268*, 17564-70.

Lee, J., and Jayaram, M. (1995a). Role of partner homology in DNA recombination. Complementary base pairing orients the 5'-hydroxyl for strand joining during Flp site-specific recombination. *J Biol Chem 270*, 4042-52.

Lee, J., Whang, I., and Jayaram, M. (1996). Assembly and orientation of Flp recombinase active sites on two-, three- and four-armed DNA substrates: implications for a recombination mechanism. *J Mol Biol 257*, 532-549.

Lee, J., Whang, I., Lee, J., and Jayaram, M. (1994). Directed protein replacement in recombination full sites reveals trans-horizontal DNA cleavage by Flp recombinase. *Embo J 13*, 5346-54.

Lin, D. M., Auld, V. J., and Goodman, C. S. (1995). Targeted neuronal cell ablation in the Drosophila embryo: pathfinding by follower growth cones in the absence of pioneers. *Neuron 14*, 707-15.

Logie, C., and Stewart, A. F. (1995). Ligand-regulated site-specific recombination. *Proc Natl Acad Sci U S A 92*, 5940-4.

Lyznik, L. A., Hirayama, L., Rao, K. V., Abad, A., and Hodges, T. K. (1995). Heat-inducible expression of FLP gene in maize cells. *Plant J 8*, 177-86.

Lyznik, L. A., Mitchell, J. C., Hirayama, L., and Hodges, T. K. (1993). Activity of yeast FLP recombinase in maize and rice protoplasts. *Nucleic Acids Res 21*, 969-75.

Mack, A., Sauer, B., Abremski, K., and Hoess, R. (1992). Stoichiometry of the Cre recombinase bound to the lox recombining site. *Nucleic Acids Res 20*, 4451-5.

Maeser, S., and Kahmann, R. (1991). The Gin recombinase of phage Mu can catalyse site-specific recombination in plant protoplasts. *Mol Gen Genet 230*, 170-6.

Marko, J. a. S., E. (1995). Statistical mechanics of supercoiled DNA. *Phys. Rev. Ser. E. 52*, 2912-2938.

McClintock, B. (1951). Chromosome organisation and genic expression. *Cold Spring Harbour Symp. Quant. Biol. 16*, 13 - 47.

McKee, B. D., Ren, X., and Hong, C. (1996). A recA-like gene in Drosophila melanogaster that is expressed at high levels in female but not male meiotic tissues. *Chromosoma 104*, 479-88.

Metzger, D., Clifford, J., Chiba, H., and Chambon, P. (1995). Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase. *Proc Natl Acad Sci U S A 92*, 6991-5.

Meyer Leon, L., Gates, C. A., Attwood, J. M., Wood, E. A., and Cox, M. M. (1987). Purification of the FLP site-specific recombinase by affinity chromatography and re-examination of basic properties of the system. *Nucleic Acids Res 15*, 6469-88.

Meyer Leon, L., Huang, L. C., Umlauf, S. W., Cox, M. M., and Inman, R. B. (1988). Holliday intermediates and reaction by-products in FLP protein-promoted site-specific recombination. *Mol Cell Biol 8*, 3784-96.

Meyer Leon, L., Inman, R. B., and Cox, M. M. (1990). Characterization of Holliday structures in FLP protein-promoted site-specific recombination. *Mol Cell Biol 10*, 235-42.

Meyer Leon, L., Senecoff, J. F., Bruckner, R. C., and Cox, M. M. (1984). Site-specific genetic recombination promoted by the FLP protein of the yeast 2-micron plasmid in vitro. *Cold Spring Harb Symp Quant Biol 49*, 797-804.

Morgan, T. H. (1911). An attempt to analyze the constitution of the chromososmes on the basis of sex- limited inheritance in *Drosophila. J. Exp. Zool. 11*, 365 - 414.

Murray, J. A., Cesareni, G., and Argos, P. (1988). Unexpected divergence and molecular coevolution in yeast plasmids. *J Mol Biol 200*, 601-7.

Nash, H. A. (1975). Integrative recombination of bacteriophage lambda DNA *in vitro. Proc. Natl. Acad. Sci. USA 72*, 1072 - 1076.

Nunes Duby, S. E., Azaro, M. A., and Landy, A. (1995). Swapping DNA strands and sensing homology without branch migration in lambda site-specific recombination. *Curr Biol 5*, 139-48.

Nunes Duby, S. E., Matsumoto, L., and Landy, A. (1989). Half-att site substrates reveal the homology independence and minimal protein requirements for productive synapsis in lambda excisive recombination. *Cell 59*, 197-206.

Nunes Duby, S. E., Matsumoto, L., and Landy, A. (1987). Site-specific recombination intermediates trapped with suicide substrates. *Cell 50*, 779-88.

Nunes Duby, S. E., Tirumalai, R. S., Dorgai, L., Yagil, E., Weisberg, R. A., and Landy, A. (1994). Lambda integrase cleaves DNA in cis. *Embo J 13*, 4421-30.

O'Gorman, S., Fox, D. T., and Wahl, G. M. (1991). Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science 251*, 1351-5.

Oettinger, M. A. (1996). Cutting apart V(D)J recombination. *Curr Opin Genet Dev 6*, 141-5.

Orban, P. C., Chui, D., and Marth, J. D. (1992). Tissue- and site-specific DNA recombination in transgenic mice. *Proc Natl Acad Sci U S A 89*, 6861-5.

Osborne, B. I., Wirtz, U., and Baker, B. (1995). A system for insertional mutagenesis and chromosomal rearrangement using the Ds transposon and Cre-iox. *Plant J 7*, 687-701.

Pan, G., and Sadowski, P. D. (1993). Identification of the functional domains of the FLP recombinase. Separation of the nonspecific and specific DNA-binding, cleavage, and ligation domains. *J Biol Chem 268*, 22546-51.

Pan, H., Clary, D., and Sadowski, P. D. (1991). Identification of the DNA-binding domain of the FLP recombinase. *J Biol Chem 266*, 11347-54.

Panigrahi, G. B., and Sadowski, P. D. (1994). Interaction of the NH2- and COOH-terminal domains of the FLP recombinase with the FLP recognition target sequence. *J Biol Chem 269*, 10940-5.

Pargellis, C. A., Nunes Duby, S. E., de Vargas, L. M., and Landy, A. (1988). Suicide recombination substrates yield covalent lambda integrase-DNA complexes and lead to identification of the active site tyrosine. *J Biol Chem 263*, 7678-85.

Parsons, R. L., Evans, B. R., Zheng, L., and Jayaram, M. (1990). Functional analysis of Arg-308 mutants of Flp recombinase. Possible role of Arg-308 in coupling substrate binding to catalysis. *J Biol Chem 265*, 4527-33.

Parsons, R. L., Prasad, P. V., Harshey, R. M., and Jayaram, M. (1988). Step-arrest mutants of FLP recombinase: implications for the catalytic mechanism of DNA recombination. *Mol Cell Biol 8*, 3303-10.

Pichel, J. G., Lakso, M., and Westphal, H. (1993). Timing of SV40 oncogene activation by site-specific recombination determines subsequent tumor progression during murine lens development. *Oncogene 8*, 3333-42.

Qian, X. H., and Cox, M. M. (1995). Asymmetry in active complexes of FLP recombinase. *Genes Dev 9*, 2053-64.

Qian, X. H., Inman, R. B., and Cox, M. M. (1990). Protein-based asymmetry and protein-protein interactions in FLP recombinase-mediated site-specific recombination. *J Biol Chem 265*, 21779-88.

Rajewsky, K., Gu, H., Kuhn, R., Betz, U. A., Muller, W., Roes, J., and Schwenk, F. (1996). Conditional gene targeting. *J Clin Invest 98*, S51-3.

Ramirez Solis, R., Liu, P., and Bradley, A. (1995). Chromosome engineering in mice. *Nature 378*, 720-4.

Rippe, K., von Hippel, P. H and Langowski, J. (1995). Action at a distance: DNA looping and initiation of transcription. *Trends. Biochem. Sci. 20*, 500-506.

Rossant, J., and Nagy, A. (1995). Genome engineering: the new mouse genetics. *Nat Med 1*, 592-4.

Rost, B., and Sander, C. (1993). Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol 232*, 584-99.

Rubin, G. M., and Spradling, A. C. (1982). Genetic transformation of Drosophila with transposable element vectors. *Science 218*, 348-53.

Sadowski, P. D. (1993). Site-specific genetic recombination: hops, flips, and flops. *Faseb J 7*, 760-7.

Sadowski, P. D. (1995). The Flp recombinase of the 2-microns plasmid of Saccharomyces cerevisiae. *Prog Nucleic Acid Res Mol Biol 51*, 53-91.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular Cloning, a Laboratory Manual Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sauer, B. (1992). Identification of cryptic lox sites in the yeast genome by selection for Cre-mediated chromosome translocations that confer multiple drug resistance. *J Mol Biol 223*, 911-28.

Schlake, T., and Bode, J. (1994). Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. *Biochemistry 33*, 12746-51.

Schmitt, W., Odenbreit, S., Heuermann, D., and Haas, R. (1995). Cloning of the Helicobacter pylori recA gene and functional characterization of its product. *Mol Gen Genet 248*, 563-72.

Schwartz, C. J., and Sadowski, P. D. (1990). FLP protein of 2 mu circle plasmid of yeast induces multiple bends in the FLP recognition target site. *J Mol Biol 216*, 289-98.

Seibler, J., and Bode, J. (1997). Double-reciprocal crossover mediated by FLP-recombinase: a concept and an assay. *Biochemistry 36*, 1740-7.

Senear, D. F., and Brenowitz, M. (1991). Determination of binding constants for cooperative site-specific protein-DNA interactions using the gel mobility-shift assay. *J Biol Chem 266*, 13661-71.

Senecoff, J. F., Bruckner, R. C., and Cox, M. M. (1985). The FLP recombinase of the yeast 2-micron plasmid: characterization of its recombination site. *Proc Natl Acad Sci U S A 82*, 7270-4.

Senecoff, J. F., and Cox, M. M. (1986). Directionality in FLP protein-promoted site-specific recombination is mediated by DNA-DNA pairing. *J Biol Chem 261*, 7380-6.

Senecoff, J. F., Rossmeissl, P. J., and Cox, M. M. (1988). DNA recognition by the FLP recombinase of the yeast 2 mu plasmid. A mutational analysis of the FLP binding site. *J Mol Biol 201*, 405-21.

Serre, M. C., Zheng, L., and Jayaram, M. (1993). DNA splicing by an active site mutant of Flp recombinase. Possible catalytic cooperativity between the inactive protein and its DNA substrate. *J Biol Chem 268*, 455-63.

Shaikh, A. C., and Sadowski, P. D. (1997). The Cre recombinase cleaves the lox site in trans. *J Biol Chem 272*, 5695-702.

Sherratt, D. (1991). Jumping Genes. *Current Biology 1*, 192 - 194.

Sherratt, D. J., Arciszewska, L. K., Blakely, G., Colloms, S., Grant, K., Leslie, N., and McCulloch, R. (1995). Site-specific recombination and circular chromosome segregation. *Philos Trans R Soc Lond B Biol Sci 347*, 37-42.

Siegal, M. L., and Hartl, D. L. (1996). Transgene Coplacement and high efficiency site-specific recombination with the Cre/loxP system in Drosophila. *Genetics 144*, 715-26.

Stark, W. M., Boocock, M. R., and Sherratt, D. J. (1992). Catalysis by site-specific recombinases [published erratum appears in Trends Genet 1993 Feb;9(2):45]. *Trends Genet 8*, 432-9.

Stark, W. M., Grindley, N. D., Hatfull, G. F., and Boocock, M. R. (1991). Resolvase-catalysed reactions between res sites differing in the central dinucleotide of subsite I. *Embo J 10*, 3541-8.

Sternberg, N., and Hamilton, D. (1981). Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. *J Mol Biol 150*, 467-86.

Sternberg, N., Hamilton, D., and Hoess, R. (1981). Bacteriophage P1 site-specific recombination. II. Recombination between loxP and the bacterial chromosome. *J Mol Biol 150*, 487-507.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res 22*, 4673-80.

Umlauf, S. W., and Cox, M. M. (1988). The functional significance of DNA sequence structure in a site-specific genetic recombination reaction. *Embo J 7*, 1845-52.

Utatsu, I., Sakamoto, S., Imura, T., and Toh e, A. (1987). Yeast plasmids resembling 2 micron DNA: regional similarities and diversities at the molecular level. *J Bacteriol 169*, 5537-45.

Van Sluys, M. A., Tempe, J., and Fedoroff, N. (1987). Studies on the introduction and mobility of the maize Activator element in Arabidopsis thaliana and Daucus carota. *Embo J 6*, 3881-9.

Waite, L. L., and Cox, M. M. (1995). A protein dissociation step limits turnover in FLP recombinase-mediated site-specific recombination. *J Biol Chem 270*, 23409-14.

Wang, Y., Krushel, L. A., and Edelman, G. M. (1996). Targeted DNA recombination in vivo using an adenovirus carrying the cre recombinase gene. *Proc Natl Acad Sci U S A 93*, 3932-6.

Wasserman, S. A., Dungan, J.M. and Cozzarelli, N.R. (1985). Discovery of a predicted DNA knot substantiates a model for site specific recombination. *Science 229*, 171-174.

Wasserman, S. A. a. C., N.R. (1985). Determination of the stereostructure of the product of Tn3 resolvase by a general method. *Proc Natl Acad Sci USA 82*, 1079-83.

West, S. C. (1994). The processing of recombination intermediates: mechanistic insights from studies of bacterial proteins. *Cell 76*, 9-15.

Whang, I., Lee, J., and Jayaram, M. (1994). Active-site assembly and mode of DNA cleavage by Flp recombinase during full-site recombination. *Mol Cell Biol 14*, 7492-8.

Wierzbicki, A., Kendall, M., Abremski, K., and Hoess, R. (1987). A mutational analysis of the bacteriophage P1 recombinase Cre. *J Mol Biol 195*, 785-94.

Yagil, E., Dorgai, L., and Weisberg, R. A. (1995). Identifying determinants of recombination specificity: construction and characterization of chimeric bacteriophage integrases. *J Mol Biol 252*, 163-77.

Zhang, Y., Riesterer, C., Ayrall, A. M., Sablitzky, F., Littlewood, T. D., and Reth, M. (1996). Inducible site-directed recombination in mouse embryonic stem cells. *Nucleic Acids Res 24*, 543-8.

Zhu, X. D., Pan, G., Luetke, K., and Sadowski, P. D. (1995). Homology requirements for ligation and strand exchange by the FLP recombinase. *J Biol Chem 270*, 11646-53.

Zhu, X. D., and Sadowski, P. D. (1995). Cleavage-dependent ligation by the FLP recombinase. Characterization of a mutant FLP protein with an alteration in a catalytic amino acid. *J Biol Chem 270*, 23044-54.

Acknowledgements

First of all, I would like to thank Francis Stewart, for, amongst other things, giving me the opportunity to do this work in his lab, for always bouncing back, for the concept of the "quiet space", and for the gift of Sam the cat.

Thanks also to Matthias Hentze and Anne Ephrussi who were on my thesis committee. I thank H. Fukuhara for the gift of plasmid pKWS1, and Michael Cox for the gift of FLP recombinase. Thanks also to the Photographic lab and the Oligo service at the EMBL, for excellent and prompt services.

Special thanks go to Valere Lounnas, who did the mathematical modelling, who taught me the difference between one end of a differential equation and the other, and who mixes a mean cocktail. Thanks also to Lutz Ehrlich, who also contributed to the kinetic project. A truly enlightening and enjoyable collaboration!

Thanks to P.O Angrand, who made the excision substrates for FLP , Kw and Cre recombinases, and who demonstrated the activity of Kw recombinase in mammalian cell culture. Special thanks also to Frank Buchholz, for the purification of Cre recombinase, and for introducing me to the worldwide football scene! Special thanks go to Sophie Chabanis, not only for giving birth to the FRED family of plasmids, but also for all those other rare qualities. A big thank you to Rein Aasland, for his infectious enthusiasm for discussion, and his unfailing support and encouragement (not to mention the fish). Thanks also to Jeanet Rientjes, for, amongst many other things, helping me to search for the DNA binding domain of FLP recombinase on the beach on the island of Elba. (Funnily enough, we didn't find it there, which is why this thesis is about something else).

Thanks also to other past and present members of the Stewart Lab: Mark Nichols, Hille Tekotte, Helen Thomas, Catherine Woodroofe, Frank Van Der Hoeven and Youming Zhang, for help and discussions over the years (or weeks). Thanks also to Toby Gibson, Phil Mitchell, Shelley Meredyth, Gisela Krauss and many others for helpful and enjoyable discussions.

Well, its 4.30 in the morning, and many more people deserve a mention here. So lets start with my parents, Mum and Dad, (thank you for having me). Then Kate and Sophie, thanks for the visits and the window to the outside. A special mention for Dave Rees, thanks for the regular reminders of the importance of excess. Thanks for all the luxury items over the years, including your marvellous company. I would also like to thank Ann Mari Voie, with whom I grew up, and Caryl Mayes, with whom I was often childish. Thanks also to Lisa Kneipp, who taught me the difference between Maggi and real Semmel Knödel. Isabelle Joncour and Jonathon Ferreira are also thanked, not only for their excellent company, but also for their special understanding of the nature of the "fin de these". Thanks to Luis Vacs, for, amongst other things, keeping my cars temporarily on the road.

When searching for inspiration in other peoples theses, as to how I should best round off my acknowledgements, I noticed that the last paragraph is usually all about the wonderful "other half". The acknowledgements range from the sublime: "With whom I explored the heights and depths of my very soul"; to the ridiculous: "My Hero!" to the realistic: "who helped me to stick in the figures". Well, in the absence of such a person, I would like to dedicate a little space here in this last paragraph to say thank you to all those people who have shared my bed and my bath over the past few years, and especially, I would like to thank: Tom Waits, William Gibson, Louis Armstrong, Patricia Highsmith, Salman Rushdie, Annie Lennox, Dick Francis, Ella Fitzgerald, Iain Banks and the London Philarmonic Orchestra. Oh, and Sam the cat.